

RESEARCH IN PSYCHOTHERAPY



RESEARCH IN PSYCHOTHERAPY

RESEARCH IN PSYCHOTHERAPY

Proceedings of a Conference,
Washington, D. C., April 9-12, 1958

Eli A. Rubinstein
Morris B. Parloff, *Editors*

This conference, financed by a grant (M-2031) from the National Institute of Mental Health, U. S. Public Health Service, was held under the auspices of the Division of Clinical Psychology, American Psychological Association, with planning and programming by an Ad Hoc Committee of the Division of Clinical Psychology; Frank Auld, Jr., Morris B. Parloff, Benjamin Pasamanick, George Saslow, Julius Seeman, and Eli A. Rubinstein, Chairman.

Copyright 1959 by the
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth St., N W.
Washington, D. C.

All rights reserved

No part of this book may be reproduced
in any form without permission
in writing from the publisher

Library of Congress Catalog Card Number 59-9192
PRINTED BY THE NATIONAL PUBLISHING CO., WASHINGTON, D. C.

Table of Contents

	Page
Contents	iii
Preface ...	v
Members of the Conference.....	vii

APRIL 9, 1958 *Dinner Session*

Goals of psychotherapy, <i>John C. Whitehorn, M.D.</i>	1
--------------------------------------------------------------	---

APRIL 10, 1958 *Morning Session*

PROBLEMS OF CONTROLS

Julian B. Rotter, Ph.D., Moderator

Problems of controls in psychotherapy as exemplified by the psychotherapy research project of the Phipps Psychiatric Clinic, <i>Jerome D. Frank, M.D.</i>	10
The research strategy and tactics of the psychotherapy research project of The Menninger Foundation and the problem of controls, <i>Lewis L. Robbins, M.D., and Robert S. Wallerstein, M.D.</i>	27
Discussant: <i>John M. Butler, Ph.D.</i>	44
Group Discussion	49

APRIL 10, 1958 *Afternoon Session*

METHODS FOR ASSESSMENT OF CHANGE (A)

Robert A. Cohen, M.D., Moderator

The dimensions and a measure of the process of psychotherapy: a system for the analysis of the content of clinical evaluations and patient-therapist verbalizations, <i>Timothy Leary, Ph.D., and Merton Gill, M.D.</i>	62
A tentative scale for the measurement of process in psychotherapy, <i>Carl R. Rogers, Ph.D.</i>	96
Discussant: <i>David Shakow, Ph.D.</i>	108
Group Discussion	116

APRIL 11, 1958 *Morning Session*

METHODS FOR ASSESSMENT OF CHANGE (B)

J. McV. Hunt, Ph.D., Moderator

A technique for studying changes in interview behavior, <i>George Saslow, M.D., and Joseph D. Matarazzo, Ph.D.</i>	125
Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome, <i>John I. Lacey, Ph.D.</i>	160
Discussant: <i>Milton Greenblatt, M.D.</i>	209
Group Discussion	221

APRIL 11, 1958 *Afternoon Session*

THERAPIST-PATIENT RELATIONSHIP

Benjamin Pasamanick, M.D., Moderator

	Page
Inside the therapeutic hour, <i>Edward S. Bordin, Ph.D.</i>	235
Some investigations of relationship in psychotherapy, <i>William U. Snyder, Ph.D.</i>	247
Discussant: <i>Maurice Lorr, Ph.D.</i>	260
Group Discussion	264

SUMMARY

Research problems in psychotherapy, <i>Morris B. Parloff, Ph.D., and Eli A. Rubinstein, Ph.D.</i>	276
-------------------------------------------------------------------------------------------------------------	-----

Preface

The content of this volume represents the proceedings of a conference held in Washington, D. C., April 9-12, 1958, under the sponsorship of the Division of Clinical Psychology of the American Psychological Association. Financing for the conference came from a research grant from the National Institute of Mental Health.

The idea for this conference was developed in the Spring of 1956 when the Research Committee of the Division of Clinical Psychology presented a recommendation, based upon a suggestion of committee member, Dr. Aaron Nadel, that the Division consider sponsoring a national conference on "objective evaluation of psychotherapy." It was further recommended that an Ad Hoc Committee be appointed, composed of both psychiatrists and psychologists, to carry out the planning necessary to obtain financial support for the proposed conference and to develop the format for the meeting.

A later and more detailed recommendation pointed out that a conference of this nature could have multiple benefits. Most importantly, this conference could afford an opportunity for taking stock of the present status of research in psychotherapy and thus provide information for and stimulus to further research. In addition, the sponsorship of such a national meeting by the Division of Clinical Psychology would demonstrate its active interest in this area of research. And, finally, by including representatives of psychology and psychiatry on both the membership of the Ad Hoc Committee and among the participants of the conference it was hoped to strengthen the research collaboration and interprofessional relations of the two disciplines.

The recommendation was approved by the Executive Committee of the Division and an Ad Hoc Committee, consisting of Drs. Frank Auld, Morris B. Parloff, Ben Pasamanick, Eli A. Rubinstein, George Saslow and Julius Seeman developed plans for the conference.

Early in June, 1957 a formal proposal for financial support of the conference was approved by the Board of Directors of the American Psychological Association and was submitted to the National Institute of Mental Health, which approved a grant of \$13,685 for the holding of the conference and for publication of the proceedings. Dr. George Kelly, as President of the Division of Clinical Psychology, served as principal investigator, while the original Ad Hoc Committee continued as a planning committee for further organization of the conference.

In this further development of the format for the conference the planning committee was much concerned with the primary audience which the conference was to serve. Was this to be purely for the invited guests—a working conference with an agenda for discussion but no formal presentations? Or were the participants to develop material for a much wider audience by preparing formal papers which would provide the substance of a published symposium volume? The conference ended up a compromise of these two alternate goals, with half the time devoted to formal presentations and half to group discussions.

Another major consideration was the focus of the conference. In the original prospectus it was stated that the general aim was to provide a series of presentations and discussions on the broad issues in research on the effects of psycho-

therapy. These were described under three major headings: A. Goals of therapy; B. Measures of change and outcome; C. Problems of controls. As will be noted from the actual program of the conference, the scope was broadened considerably from an emphasis on effects of psychotherapy to include also various aspects of process research.

As indicated in the Table of Contents, the conference was organized into four topic sessions, each occupying half a day. The eight formal papers were all prepared in advance and distributed to the conferees before the meeting. The speakers then gave half-hour summaries of their papers, with each discussant presenting a formal discussion paper, also prepared in advance, in response to the two addresses for that topic session. The last half of the time for each topic was devoted to group discussion among all participants. There were no formal subgroup meetings but luncheon and dinner conversations plus social hours during the two evenings produced the equivalent of a number of informal small group meetings. On the evening preceding these two full days Dr. Whitehorn presented his opening address on "Goals of Therapy." And, on the last morning following these two days, there was a closing session of three hours at which time the group considered and discussed in turn each of the four topic sessions.

A number of people whose names do not otherwise appear in this volume were helpful in the development of various phases of the conference planning. Dr. James E. Birren of the National Institute of Mental Health, Dr. Leland Bradford of the National Training Laboratory, and Dr. Jonathan Cole of the National Institute of Mental Health, provided valuable suggestions on the administration of various phases of the conference format.

Acknowledgment is due Dr. Arthur Hoffman of the American Psychological Association for assistance in the many details necessary to publishing this volume. Dr. Roger Russell, Executive Secretary of the American Psychological Association served as Financial Officer for the grant. It is appropriate, not only to acknowledge the financial assistance of the National Institute of Mental Health, but to express appreciation to Mr. Philip Sapir, Chief, Research Grants and Fellowship Branch, for his consultive guidance in regard to the grant application.

It would ordinarily be superfluous and inappropriate in a volume of this nature to acknowledge the role of the conference participants since the content is essentially their own product. However, in the preparation of this book, the proceedings have been somewhat edited and the contributions of some of the participants are not as apparent as they were at the conference itself. The moderators of each of the sessions, Drs. Julian Rotter, Robert A. Cohen, J. McV. Hunt, and Benjamin Pasamanick for the four topic sessions, and Dr. George Saslow for the final morning summary session, each provided the necessary guidance to keep the discussion pertinent and directional. An especial acknowledgment is due Drs. David Hamburg, Rosalind Dymond Cartwright, Joseph D. Matarazzo, and Hans Strupp. On the last morning, each of these participants provided a summary presentation of one of the topic sessions, including their own interpretations of the themes of that session's discussion period. These summary presentations then served as the basis for the discussion period of the last morning.

ELI A. RUBINSTEIN
MORRIS B. PARLOFF, Editors

Participants and Guests of the Conference on Research in Psychotherapy

CONFERENCE PARTICIPANTS

- FRANK BARRON, Visiting Professor, Wesleyan University, Middletown, Connecticut
- EDWARD S. BORDIN, Professor of Psychology, University of Michigan, Ann Arbor, Michigan
- JOHN M. BUTLER, Executive Secretary, The Counseling Center, University of Chicago, Chicago, Illinois
- DESMOND S. CARTWRIGHT, Assistant Professor of Psychology, University of Chicago, Chicago, Illinois
- ROSALIND DYMOND CARTWRIGHT, Research Associate (Associate Professor), The Counseling Center, University of Chicago, Chicago, Illinois
- ROBERT A. COHEN, Director of Clinical Investigations, National Institute of Mental Health, Bethesda, Maryland
- ALLEN T. DITTMANN, Clinical Psychologist, Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland
- JEROME D. FRANK, Associate Professor of Psychiatry, Johns Hopkins Medical School, Baltimore, Maryland
- MILTON GREENBLATT, Assistant Superintendent and Director of Research and Laboratories, Massachusetts Mental Health Center, Boston, Massachusetts
- DAVID A. HAMBURG, Chief, Adult Psychiatry Branch, National Institute of Mental Health, Bethesda, Maryland
- J. McV. HUNT, Professor of Psychology, University of Illinois, Urbana, Illinois
- KENNETH S. ISAACS, Research Associate, University of Illinois, College of Medicine, Chicago, Illinois
- JOHN I. LACEY, Chairman, Department of Psychophysiology and Neurophysiology, Fels Research Institute, Antioch College, Yellow Springs, Ohio
- TIMOTHY LEARY, Director, Kaiser Foundation Psychology Research, Oakland, California
- MAURICE LORR, Chief, Neuropsychiatric Research Laboratory, Veterans Benefits Office, Washington, D. C.
- LESTER LUBORSKY, Senior Psychologist, Research Department, Menninger Foundation, Topeka, Kansas
- JOSEPH D. MATARAZZO, Professor of Medical Psychology, University of Oregon Medical School, Portland, Oregon
- MORRIS B. PARLOFF, Chief, Section on Personality, Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland
- BENJAMIN PASAMANICK, Professor of Psychiatry, College of Medicine, The Ohio State University, and Director of Research, Columbus Psychiatric Institute and Hospital, Columbus, Ohio
- LEWIS L. ROBBINS, Co-Chairman, Menninger Foundation Psychotherapy Research Project, Menninger Foundation, Topeka, Kansas
- CARL R. ROGERS, Professor, Departments of Psychology and Psychiatry, University of Wisconsin, Madison, Wisconsin
- JULIAN B. ROTTER, Director, Psychological Clinic, Ohio State University, Columbus, Ohio
- ELI A. RUBINSTEIN, Program Analyst, Training Branch, National Institute of Mental Health, Bethesda, Maryland
- GEORGE SASLOW, Professor of Psychiatry, University of Oregon Medical School, Portland, Oregon
- JULIUS SEEMAN, Professor of Psychology, George Peabody College, Nashville, Tennessee
- DAVID SHAKOW, Chief, Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland
- WILLIAM U. SNYDER, Professor of Psychology; Director, Psychology Clinic, The Pennsylvania State University, University Park, Pennsylvania
- HANS H. STRUPP, Director of Psychological Services, N. C. Memorial Hospital, University of North Carolina, and Associate Professor of Psychology in Departments of Psychiatry and Psychology, Chapel Hill, North Carolina
- ROBERT S. WALLERSTEIN, Associate Director, Department of Research, Menninger Foundation, and Co-Chairman, Menninger Foundation Psychotherapy Research Project, Menninger Foundation, Topeka, Kansas

DONALD L. BURHAM, Director of Research,
Chestnut Lodge Research Institute, Rock-
ville, Maryland

MABEL B. COHEN, Editor, "Psychiatry,"
Training and Supervising Analyst, Wash-
ington Psychoanalytic Society, Inc., Wash-
ington, D. C.

H. MAX HOUTCHENS, Chief, Clinical Psy-
chology Division, Psychiatry and Neu-
rology Service, Department of Medicine
and Surgery, Veterans Administration,
Washington, D. C.

GEORGE N. RAINES, Director, Department of
Psychiatry, Georgetown University Medi-
cal School, Washington, D. C.

DAVID MCK. RIOCH, Director, Division of
Neuropsychiatry, Walter Reed Army In-
stitute of Research, Washington, D. C.

ROGER W. RUSSELL, Executive Secretary,
American Psychological Association,
Washington, D. C.

PHILIP SAPIR, Chief, Research Grants and
Fellowships Branch, National Institute of
Mental Health, Bethesda, Maryland

JOHN C. WHITEHORN, Psychiatrist-in-Chief,
The Johns Hopkins Hospital, Baltimore,
Maryland

Goals of Psychotherapy

JOHN C. WHITEHORN, M.D.

On the eve of a conference in which we may expect vigorous assertions and the impact of hard facts upon fondly-held conceptions, your program committee has wisely provided this occasion for agreeable companionship in the pleasures of dining. In breaking bread together we have followed a time-honored custom for establishing rapport.

Now your chairman has called upon me to offer you my thoughts on the topic "Goals of Psychotherapy." For the purpose of this conference, this is a central—and crucial—topic. Earnest and scholarly minds are devoting careful thought to the task of elucidating the processes involved in psychotherapy, with much attention to discriminating observation and experiment. This conference has been called for the purpose of considering concepts and methods for the study of psychotherapy. For the rational study of any human activity it is necessary to know or to postulate the purpose or goal of that activity. The question, "What is the goal of psychotherapy?" is therefore the key-question of this conference—the key issue from which all other issues derive their significance.

You will readily understand, therefore, that I felt much complimented when asked by your program committee to speak on this key issue. Yet I experienced also some hesitation and indecision in accepting the invitation; in part because of my awareness that the inherent difficulties of this field are compounded by a swarm of aggravating additional difficulties created by professional rivalries and by human propensities to take sides, to establish group loyalties, and to struggle over whose views shall prevail.

I did accept the task, and I shall now proceed to it.

First, I can offer the formula: "The goal of psychotherapy is health." To many, this formula will appear self-evident, axiomatic,—indeed, so self-evident that one runs much risk of looking ridiculous even to consider it a serious topic for discussion. Yet some discussion may be justified, if merely for clarification.

"Health" is a broad term, of multiple meanings. It needs definition and specification. To those in command of a task force of workers, the health of the workers means their ability to work effectively. To the person himself, it means something more—it means "to enjoy health"—not merely to function appropriately and effectively, in work or in play, or in mere existence, but to do so with that inner sense of comfort, pride and satisfaction, which may be called the glow of health. Since human beings are unalterably forward-looking, it means also pleasurable *expectations* of continued competence and satisfaction in functioning. To work good, to feel good, to expect good—these might be called the primary dimensions of health. (When I say good, I mean well.)

If one attempts, for study purposes, to construct a scale of health, these "good" terms seem, however, a bit too vague. One seeks, and finds, more readily scalable positions in the negative directions of these dimensions—in degrees of disability, in degrees of distress, in degrees of dread.

I have delineated, thus, very briefly a frame of reference, and a statement of the goals of psychotherapy which in my judgment has much practical value. The

statement seems to get to the heart of the matter, and in a manner which lends itself to actual research operations. I feel fairly well fortified in this opinion by my acquaintance with the investigations of Dr. Jerome Frank, and his collaborators, who have made much use of practical scales for two of these three dimensions.

Let us, however, give another look at the statement that the goal of psychotherapy is *health*. By this choice of words I have brought this discussion into a region of possible controversy. The use of the word "health" has implied that the territory we discuss—namely psychotherapy—is a province of the healing profession, namely the medical profession. This semantic gesture has not logically done any more than is done by the word "therapy"—which also allocates the activity to the healing profession—yet it is possible to pronounce the word "psychotherapy" with such emphasis upon the first part as to imply that it is the special province of the psychologist. Yet I wonder. We have also a term "pharmacotherapy"—(meaning healing by drugs) yet we do not presume—no matter how we pronounce the word—that the pharmacist takes the responsibility for the patient's care.

In this way I illustrate what I meant by my earlier comment that the inherent problems of this field may be compounded by human propensities to take sides, to establish group loyalties and to struggle over whose views shall prevail.

Mr. Chairman: I trust that you will not think I have deliberately chosen my words to stir up this controversy for the sheer pleasure of controversy. I could not pretend to be without convictions or bias. I believe that psychotherapy is, primarily, a responsibility of the medical profession. I believe also that the medical profession needs all the assistance it can get from psychologists, and others, in scientific investigation aimed at the understanding and the mastery of the

processes and procedures involved in psychotherapy. Furthermore, it appears to me that the processes involved in psychotherapy have close affinity, and even identity, with processes involved in personal counselling, in legal advice, in religious leadership and in many other forms of interpersonal transactions, wherein the primary responsibility does not lie with the medical profession. Whatsoever can be done in any of these fields to gain a better insight and a surer knowledge of such processes brings a potential contribution to the improvement of psychotherapy.

I have made this brief digression to signalize to you my awareness that controversial connotations may be aroused by the word "health" and by other heavily weighted words that I will have to use. I hope by this one show of transparent honesty to dispel suspicions that I may be setting verbal traps for any one.

To recapitulate, now, how far we have gotten, up to this point: As a first approximation, I offered the formula that the goal of psychotherapy is *health*. Then I stated that practical specifications, for working purposes, are most conveniently located in the negative directions of the three dimensions of health, namely in degrees of disability, in degrees of distress, and in degrees of dread. Perhaps I can borrow a phrase from Hollywood and call this the "3-D" system of evaluating progress toward the goals of therapy.

I am sure you will recognize that this "3-D" system is contrived for easy remembering—*Disability, Distress and Dread*—and is designed to take advantage of our mental habit of thinking in three dimensions.

Perhaps it is also merely a *habit* of the medical mind that makes it more convenient for us to think along the negative directions of these dimensions. Symptoms of illness are the regular medical stock-in-trade; we are accustomed to

look for, to recognize and to pigeon-hole the symptoms of illness. We are not in the habit of thinking about the symptoms of health, and have not developed habitual pigeon-holes for that purpose. The good physician does not really remain completely oblivious to the symptoms of health. He observes the brighter eye, the more erect posture, the rising cadence of speech, etc., etc., but his system of notation or record-keeping is so strongly biased toward the pathological that he even writes down the word "negative" when he means healthy. Such is the power of jargon over thought.

The medical approach has also given us disease categories in the areas of our concern in this conference, for example, Hysteria, Schizophrenia, Manic-Depressive Psychosis, Obsessive Neurosis, etc., etc. In general medicine there have been magnificent gains achieved by this mode of guiding the search for therapy, whereby the disease is made so to speak the target of a specific therapy, or of specific measures for control or prevention. Diabetes, Syphilis and Tuberculosis may be cited as examples of fair degrees of success in target-oriented therapy; Diphtheria and Malaria as examples of even more brilliant success; Leukemia and Lupus Erythematosus as examples which have so far defied the search for specific therapy, but offer some promise.

It is no trade secret that many psychiatrists are skeptical as to the validity of the disease-entity concept for the conditions which have been named—Hysteria, Schizophrenia, Manic-Depressive Psychosis, Obsessive Neurosis, and others, and as a corollary there is skepticism as to the likelihood of devising specific therapies. I would not favor abolishing these terms, for they have pragmatic value in facilitating communication, enabling workers to compare and collate observations, with a moderate measure of probability that they are talking about the same types of patients. The usual skept-

tic's position is that these categorizing terms indicate, at the descriptive level, more or less distinguishable constellations of symptoms, with some likelihood but no certainty that distinctive pathological processes (either mental or non-mental) may be found to characterize or maybe even to determine the named categories. This is approximately the current position of psychiatrists, as represented in the official system of classification, built around the concept of Reaction Patterns or Reaction Types. At this point, opinions begin to divide, some feeling much more convinced than others as to the existence and characteristics of specific pathogenic processes involved in specific reaction types. Some of the current hypotheses employ biochemical models of thought, some employ psychological models. It is from the latter—the psychological models—that one derives hypotheses for specifically disease-target-directed psychotherapeutic strategy.

If one postulates more or less specific intrapsychic conflicts as the pathogenic processes causative of more or less specific morbid reaction types or mental diseases, then the goal of psychotherapy can be fairly definitely specified as being: to reverse the pathogenic process by resolving the intrapsychic conflict. For convenience of discussion it may be convenient to use the expression "the target of psychotherapy" rather than the "goal of psychotherapy" when the effort is directed *against* the disease process.

Then, it may also be postulated that one of the crucial factors in the pathogenicity of an assumed intrapsychic conflict lies in the unconscious nature of the conflict, and, on this assumption one of the goals or targets of psychotherapy would be to develop conscious awareness of the conflict, as a necessary step toward its resolution.

On such a two-factor theory of causation, it is conceivable that *one* or the *other* of the factors may be made the

primary target of therapy, that is: A, plans might be directed first to making the unconscious conscious, whereby the rational resources of the conscious mind might be enlisted with clear awareness, in the effort to resolve the conflict, or, B, plans might be directed toward providing for some "corrective emotional experiences" for the patient whereby some resolution, or more tolerable modification of the conflict might first be gained, without clearly spot-lighting the issue in consciousness. If such a therapeutic aim as outlined under B proved attainable, working essentially at non-rational or unconscious levels, the pathogenicity of the presumed conflict might be sufficiently modified or it might even, conceivably, be resolved, to such an extent that relief or even recovery might be gained. A more complex strategy might be conceived of, combining A and B, whereby corrective emotional experiences at unconscious levels, might modify the conflict enough to enable the patient to bring it more readily into consciousness.

For many who seek their guide-lines for psychotherapy in the Freudian system of concepts, or are familiar with it, these thoughts which I have just presented make sense, I believe, even though I have used rather abstract language. The abstractness of the language has not, however, guided me past all controversy. Some of the words I have used, such as "corrective emotional experience" could precipitate controversy here. One of the words which I intentionally did not use, namely the word "insight" could also precipitate controversy. But I do not choose to digress in this direction.

In these more recent paragraphs, I have been presenting a target-directed concept of psychotherapy, that is a conception of an effort directed *against* a pathogenic process. This is in the classical medical model of "the remedy for the disease." Yet there is also a long and

honorable medical tradition to the effect that the doctor treats patients, not diseases. Even in the abstract terms which I used I trust that the attentive listener caught implications that the patient was to be involved actively in the psychotherapeutic process.

At one time I said "whereby the rational resources of the conscious mind might be enlisted in the effort to resolve the conflict" and at another time I said that the patient might be enabled to bring the conflict into consciousness. In both instances I was speaking of the activities and resources of the patient, and clearly implied that the therapist would do well to evoke these resources and activities. I wish to make this point definite and explicit, that even in discussing target-directed psychotherapy aimed against disease-processes,—that is, efforts directed *against* the presumed cause—I have used language which clearly implies an evocative meaning in the therapeutic potential resides in the patient, and one of the aims of psychotherapy is to evoke this potential.

Here I wish to pursue another implication. To evoke effort from the patient is to exercise a leadership function, and leadership means leading in some direction. In what direction?

We could use here the formula I offered first ("The goal of psychotherapy is health") and say that the psychotherapist aims to evoke that in the patient which leads toward health.

For some purposes we could stop at this point, but in this situation I feel impelled to press on, and to state more explicitly an impression which I have gained in the study of successful psychotherapists, namely that they do function as leaders, and that a psychotherapist does lead in a direction which according to his values means a good life for the patient. Values vary; conceptions of the good life vary; leadership toward the

good life implies wisdom, as well as knowledge and skill. Therefore, in venturing to assert that psychotherapy involves leadership in directions going beyond the bare concept of health, I am entering territory where taste and preference prevail. I wish to avoid misunderstanding here. It is not my intention to say that this should be so; rather I merely record my impression as to what is the fact, namely that successful psychotherapy, when subjected to empirical scrutiny, is found to involve leadership toward preferred values, toward the therapist's conception of what constitutes value in life.

Specifically, some psychotherapists lead their patients toward freedom, others toward conformity. Neither goal is completely attainable for human beings. Interdependence is inherent in the institutions of domestication and socialization within which human life is lived, which require some degree of conformity; yet some measure of free play and spontaneity is necessary to make these institutions work. In the jargon of the sociologists, each must play some role or roles, and this means conforming in some measure to the expectations of others, yet the development of personal integrity, which is also a social necessity, requires that individuals have some feeling of inner freedom and self-fulfillment in enacting these roles.

Psychotherapists, as I have perceived them, have definite biases, placing high value upon one or the other of these goals,—freedom or conformity—as expressions of their own personal orientation to life, and they manifest these value preferences in their therapeutic operations. Such manifestations of preference are ordinarily tacit or implied, rather than explicit, yet I am inclined to think that these manifestations of value preference constitute one of the major therapeutic forces, evoking like desires and efforts in susceptible patients and leading

them toward the indicated goal, tempered by whatever wisdom for compromise can be brought to bear by both in the resolution of hindering conflicts, where issues had been too sharply drawn and frozen.

As an exemplar of those psychotherapists who place a high value on freedom, and who operate implicitly to evoke like aspiration and effort in their patients, I name Sigmund Freud. It happens to be true that he for a time expounded a doctrine of complete determinism. Yet he showed himself capable of a noble inconsistency in that his mode of therapeutic operations seems well designed to activate and cultivate the patient's potentiality for freedom. Freud exemplified in his own life and work the ideal of scholarly detachment and independence of thought, and his therapeutic efforts can be interpreted I think as springing from a desire to assist others toward a similar freedom and detachment. The ideal human condition toward which Freud's therapy aimed appears to me to have been a state of freedom in which a person could, through understanding, hold himself inwardly free from the coercive prohibitions of society and free also from the coerciveness of blind biological impulses.

Such an ideal state of inner freedom is more earnestly desired by some than by others. It is more likely to be an aspiration of intellectuals and upper-class persons than of lower-class persons. Many who think they aspire to freedom are only conforming to a fashion, and shrink from the reality of it. Many patients who are unsuitably encouraged to seek emotional health by striving toward the freedom goal, without wise consideration of other needs, fail in their quest, or quickly drop out.

As exemplar of those psychotherapists who place a high value on conformity, and who operate implicitly to evoke like aspiration and effort in their patients, I have no specific name. I suspect that

my restraint in not naming names here arises in part from my own personal bias in favor of freedom, which would make it seem to me like an accusation to label someone as conformity-oriented. Yet I have known many patients who were helped by psychotherapy which was basically conformity-oriented.

I suspect that you may share with me some sense of shame, as if conformity were a nasty word. Some seek a gentler term—adjustment—and, as if in embarrassment at so ignoble a profession of value, embellish their discussions with much talk about reduction of tensions, of homeostasis, of sublimation.

When the policeman admonishes his client, "Now, come along quietly; don't make any trouble, and everything will be all right," he is exemplifying, in caricature, what might be considered a pseudo-psychotherapy whose goal is a sort of universal gaol. It is not to this extremely negative type of conformity-oriented psychotherapy that I wish to direct attention now, but rather to a more positive implication of conformity from which I think it derives its real appeal and its genuine psychotherapeutic meaning.

In admitting to some feeling that there is something shameful in conformity, I think that we manifest our unwise submission to one of the prejudices of our time and of our culture. I misdoubt the presumed ignobility of conformity. Woman conforms to man, and man to woman with enthusiasm and pride; the suckling infant conforms to mother, and mother to infant, with an eagerness which gives one warrant to think that conforming behavior may be infused with an inner glow of enthusiasm quite as genuine and noble as is the aspiration for freedom. So, too, in the infinitude of human interactions, domestic and social, many transactions and relationships of sustenance and sharing may come to partake also of this inner eager-

ness for conforming. Yet it requires, for many of us, in our time, something of an effort to perceive and maintain a view of the ennobling nature of conforming impulses. The conformity impulses are so conveniently exploitable that we have had much reason to become distrustful of them. Despite this uneasy distrust lest one be fooled into becoming a tool for the crafty or selfish purposes of others, there are highly respected professions, such as medicine and nursing, in which service to the needs of others, which is unmistakably a conforming type of behavior, has become an institutionalized ideal, as well as one of the natural human motivations. The manifestations of conforming impulses are sometimes labeled "unselfishness"—an unfortunate word which makes an absurd mystery out of a quite natural impulse.

Some psychotherapists, as I have said, are so oriented to life that they are acquainted with the enthusiasm and eagerness of spirit which may be inherent in the conforming impulse; they manifest this type of action, and the inherent enthusiasm, in their transactions with others, and thus may evoke in susceptible patients some like enthusiasm. Why does this phenomenon have special relevance for neurotic or psychotic illness? Why does it count in psychotherapy? I can offer a partial answer.

As I see it, one of the common sources of inner conflict—one of the fairly general formulations for ill-suppressed resentment—lies in the fact that many persons, while enacting their roles in life (even aptly selected roles) become somewhat irked by the sense of constraint or coercion implied by the expectations of others. (One common verbal formulation of this complaint: "I'm just taken for granted around here.") The vague uneasiness over being exploited in one's role dampens one's enthusiasm and thereby enhances the vague sense of feeling cheated or belittled. Whatever helps one

to recapture some sense of dignity and worth in the devoted enactment of one's role helps to mitigate this inner conflict.

In World War II I saw soldiers who had "blown their tops" after prolonged experience of combat, having lost their meaningful human orientations in a universe apparently gone mad with killing and being killed, who in the medical station recaptured some sense of meaningful devotion, and rejoined their units with eagerness. It seemed to me that the medics accomplished this therapy, in part, and in very significant part, by their quiet enactment of a meaningful role of devoted service, as well as by their words or their drugs.

So, too, in civilian life it has frequently seemed to me that emotionally disturbed patients, irked by those no-longer-glamorized expectations of others which defined their roles, have found in nurses and doctors a manifestation of quiet dignity and pride in devoted service which gave them support and sustained them, in meeting—with less sense of irksomeness—the expectations of their roles. Perhaps some of the neurotic habitues of outpatient departments seek, in a vague and groping way, for a comparable infection to lighten the burden of unrequited devotion in their role in life.

I have spoken at some length now regarding two special value systems—that of freedom and that of conforming to the expectations of others—which constitute, as I see it, goals toward which differing types of psychotherapists tend to lead their patients,—goals going somewhat beyond a limited definition of health. Manifestations of behavior and attitude oriented toward these values appear to play a significant part in the processes of psychotherapy. These are phenomena deserving scientific study.

I have also indicated that patients differ in their aspirations, and in their susceptibility to leadership in these directions. Some respond well to the type of

psychotherapist oriented toward the value of freedom, others to the type oriented toward conformity. In very large measure we deal here with implicit goals and operations, not explicitly avowed, inherent in the transactions between patient and therapist, expressions of personality rather than planned technique.

For purposes of scientific study, however, these aspects of psychotherapy can to some extent be brought under observation and empirical scrutiny. For some years, Dr. Barbara Betz and I have been making studies of the transactions between patients and doctors, which have some relevance to this issue.

One of the difficulties in such study arises from the obvious fact that freedom-valuing therapists and conformity-valuing therapists do not exist in pure form. It is a question of the relative emphasis upon these values, in the personalities of therapists and patients, and in the inevitable practical compromises which patients have to work out in their life situations.

Mr. Chairman: In my discussion of the implicit goals of freedom and conformity, I perceive that I have strayed outside the boundaries of my primary formulation that the goal of psychotherapy is health. Insofar as I have been considering the ultra-medical question of what health is for, I have gone beyond the goal line. Insofar as I have been considering the influence of such value orientations in actual psychotherapeutic processes, I have really been discussing therapeutic operations, which is not the topic for this paper.

Coming back to the central issue of health, as the goal of psychotherapy, I think that it still requires some amplification, in the direction we usually call rehabilitation. This is an aspect of treatment sometimes lost from view when attention is focused sharply upon the disease-process as the target of therapy. In psychotherapy, however, the personal

relationship tends to arouse in the therapist a concern for the patient's rehabilitation, either in the sense of his resuming a former functional role and capacity or in assuming new roles. In general psychiatric practice, institutional or private, there has been in recent years increasing practical attention to rehabilitation as part of medical responsibility, and I believe it is justifiable to state as an historical fact that this change of professional attitude—this improvement of professional attitude—has come about because of the more widespread interest in psychotherapy and the effects of this interest in inducing a greater dedication to the patient's welfare.

Increased interest in rehabilitation as a goal of therapy has been part of a change in attitude which I might succinctly describe as a re-affirmation of the melioristic approach.

During the hey-day of psychoanalytic pretensions, its proponents viewed that method of therapy as the one and only instrument for getting at, and really curing, the cause of mental ills; but experience and comparisons have led to a more general recognition of the imperfections and limitations of psychoanalysis. During this re-evaluation, melioristic aims in psychotherapy have regained respectability. The melioristic approach may be characterized as that in which one views the patient as a person functioning not very well or very happily and seeks to help him to a better mode of functioning; as contrasted with the perfectionistic view of seeking to achieve the complete elimination of his disease, through the radical cure of its cause.

Before I close this discussion, I must manage to give brief consideration to another topic—immaturity. Certainly the therapist's manner of thinking about immaturity has much relevance for his manner of approaching his patients' problems and of conceiving the goals of

therapy, particularly for the younger patients whose personality development has been impeded but whose potentialities for further development seem good. It is necessary to have some scale or schedule of levels of immaturity for the rational discussion of this topic, and I have on other occasions discussed formulations based upon the stepwise development of domestic and social attitudes and values, which I shall not present at this time. Some systematic conception of immaturity appears necessary in dealing with the problems aptly called "unfinished business," which we so often encounter in psychotherapy. By this phrase, I refer to those vague yet resentful ruminations, and those repetitious re-enactments of issues involved originally in some long-ago conflict not adequately resolved at the time and therefore remaining to pester the patient in poorly recognized ways until the basic issue can be faced, cleared up and properly settled to the patient's satisfaction. In suitable cases one can also envisage progress from low levels of immaturity toward higher levels as a reasonable goal of therapy, and even evaluate such progress on a scale.

Now, Mr. Chairman, I shall bring this discussion to an end, with a brief recapitulation.

In simplest terms, it can be said that the goal of psychotherapy is health. For working purposes, three dimensions of health are specified—to work well, to feel well and to expect well—most conveniently scalable in the negative directions of Disability, Distress and Dread. Along these dimensions, one can construct usable measures of success or failure in therapy.

If one chooses to focus sharply upon the pathological, the aims of psychotherapy can be formulated as target-directed efforts against pathogenic processes—as exemplified in such concepts as intrapsychic conflict and the Uncon-

scious. Even when committed to such specific target-directed psychotherapeutic efforts, one needs to recognize the evocative character of the process—that therapeutic potential resides in the patient and that part of the psychotherapeutic process is to evoke this potential.

Convictions that one has a correct understanding of pathogenesis incline one toward the perfectionistic formulation of goals; skepticism and humility incline one toward the melioristic formulation of goals. In the melioristic context, rehabilitation seems logically a part of therapy.

Immaturity in personality development may be a limiting factor in setting goals for psychotherapy. A scale of immaturity can also serve for measuring progress toward maturity, when that occurs.

Beyond the goal of health, as it may be defined strictly, lies a region of value-judgments, regarding preferences as to what life and health are for. High value placed upon freedom, or upon conformity, may implicitly determine the strategic aim of the therapist. Behavioral manifestations of these value-orientations are important factors in the processes of psychotherapy.

Problems of Controls in Psychotherapy as Exemplified by the Psychotherapy Research Project of the Phipps Psychiatric Clinic

JEROME D. FRANK, M.D.

In the broadest sense, the purpose of controls is to answer the question: how sure are you that you really know what you think you know? Problems of control arise only after a researcher thinks he knows something—that is, after he has an hypothesis that certain variables are related in a certain way—and he wishes to determine whether he is right. The purpose of controls, in other words, is to exclude alternative hypotheses. The level of certainty at which the truth or falsity of an hypothesis can be established is a function of the accuracy with which the relevant variables can be identified, measured and manipulated. Therefore the degree of possible and desirable control in a particular field of study depends on its state of development. In its pre-scientific stage important insights may be achieved without the use of any controls worthy of the name. Even at this level, however, since the researcher only explores regions where he expects to find something, he is being guided by implicit hypotheses, and the use of crude controls may facilitate his search. Darwin, for example, used a kind of control when he made a special point of jotting down observed phenomena which seemed to refute his tentative hypotheses.

Research in psychotherapy attempts to set up and test hypotheses subsumed under the general question: what kinds of therapist activity produce what kinds of change in what kinds of patient. That is, the independent variables lie in the patient's state before the therapist's intervention and in the therapist's activity, the dependent variables in changes in the

patient's feelings and behavior. Since few of these variables are as yet adequately defined and the researcher can directly observe or manipulate only a few of those which are important, it is obvious that the field of psychotherapeutic research is still at a relatively primitive level. It is, however, possible to design studies which are controlled at least to some extent in that they permit planned, though crude, manipulations of certain variables and randomization of others.

For purposes of discussion of controls in psychotherapy, the division suggested by Edwards and Cronbach (6) into person variables, situation variables, and response variables seems especially convenient. The personal attributes of the therapist are situational variables from the patient's standpoint, and are therefore considered under this heading.

In order to give focus to the discussion, I shall draw chiefly from a study of psychotherapy with psychiatric outpatients carried out at the Johns Hopkins Hospital, in which we have barked our shins against most of the major problems of control.¹ The project took its origin

1. These studies were supported by the Veterans Administration and by a research grant from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service. The research staff consisted of Lester H. Gliedman, M.D., Stanley D. Imber, Ph.D., Earl H. Nash, M.S., and Anthony R. Stone, M.S.S.W. in addition to the writer. We wish to express our grateful acknowledgment to Dr. Morris B. Parloff for his crucial contributions to the planning and early phases of the project.

from the lack of demonstrable difference in improvement rates reported by proponents of different therapies (2), suggesting the likelihood that all forms of therapy, including some which are not called psychotherapy, have much in common. The task we set ourselves was to try to identify attributes of patients determining their responsiveness to these common features of psychotherapy. This required the use of more than one therapist and type of therapy and a design which would make it possible to determine if any attributes of patients were related to improvement regardless of therapist or therapy. In addition, the design permitted analysis of the data to discover possible specific contributions of different therapists and therapies to the obtained results.²

The patients were selected for the project on their initial visits to the outpatient department of the Johns Hopkins Hospital. They were white, aged 18-55, and of both sexes. Only those with organic brain disease, antisocial character disorders, alcoholism, overt psychosis, or mental deficiency were excluded. They were further characterized initially by the usual clinical diagnostic categories, by an inventory covering various aspects of their attitudes and behavior deemed relevant to therapy, and by initial scores on the scales used to measure change, described below.

With respect to the situational variables, the therapists were three members of the psychiatric resident staff in the second year of training. Each had done considerable individual therapy and had conducted one therapeutic group under supervision. Three forms of therapy were used: group, individual, and "minimal." Group and individual therapy were guided by the therapeutic philoso-

phy of the Phipps Clinic. In general, the therapist's aim is to establish a relationship with the patient which will help him identify and correct current distortions in his interpersonal perceptions and behavior. In individual therapy this implies relatively greater emphasis on the present than the past; in group therapy, emphasis on group interactions rather than on events transpiring outside. Examination of historical origins of these distortions is seen as a means of clarifying them, when appropriate, not as an end in itself. Patients received group therapy one and one-half hours once a week, individual therapy one hour a week. Minimal therapy consisted of a brief infrequent interview, not more than one-half hour every two weeks, focused on the patient's complaints and how he might best deal with them. The reasons for the choice of these three forms of treatment will be considered below.

The response variables were changes in the patient's subjective discomfort and his social ineffectiveness. These were chosen as representing the least common denominator of the aims of all the healing arts, including psychotherapy. However else a patient may change under treatment, unless he becomes more comfortable and more effective, it is hard to maintain that he has improved. Discomfort was defined in terms of forty-one symptoms or feelings which the patient reported as distressing. Ineffectiveness was defined in terms of fifteen types of behavior generally recognized as socially ineffective, rated by interviewers on the basis of information obtained from the patient and an informant. We attempted to measure decrease in discomfort and ineffectiveness rather than increase in comfort and effectiveness, because it proved much easier to define degrees of malfunctioning than of successful functioning. It is easier to define illness than health.

2. For a more detailed account of the design see Frank, Gliedman, Imber, Nash, & Stone (8).

As to the research design, each psychiatrist conducted all three forms of treatment, and patients were assigned at random to each by the research staff. Each psychiatrist treated 18 patients, six in each of the three forms of treatment. Psychiatrists and patients were urged to remain in contact for at least six months, unless the psychiatrist felt that the patient had received maximum benefit before that time.

Each patient, whether he stayed in treatment or not, was re-evaluated six months after entering therapy (or when therapy terminated if this occurred between one and six months), again six months later, and at yearly intervals thereafter. In most cases a relative or close friend of each patient was interviewed separately at approximately the same time as the patient.

One of the major problems in attempting to do controlled studies in outpatient therapy is the difficulty in carrying through an experimental design. Because of the mobility of the American population, and the fact that psychotherapy competes with so many other activities in patients' lives, attrition, missed appointments and the like create severe problems for maintaining any design which extends over time. Our design called for 54 patients to receive six months of treatment by three therapists. By starting with 91 patients and conducting the treatments over about 18 months, we finally succeeded in obtaining 54 patients, of whom 37 (68%) completed at least six months of treatment and 48 (90%) had four months or more of treatment. By expending considerable effort we obtained follow-up interviews on 53 of the 54 treated patients at one year and 48 at two years. Since then attrition has been marked.

With this example in mind, we may turn to consideration of some problems of control of patient, situational, and response variables.

CONTROL OF PATIENT VARIABLES

A problem which plagues all research with psychiatric patients is the adequate definition of the sample to be studied (25). The criteria used ideally should be communicable with sufficient clarity and precision so that other workers by using them can duplicate the sample. At the same time they must be relevant to psychotherapy. None of the customary criteria are adequate in these respects.

It is relatively easy to specify what may be termed actuarial characteristics of a research population, such as age, sex, race and so on, so that others can duplicate the sample. The relevance of many of these characteristics, however, is questionable. If to play safe, a large number of criteria are included, even though many are suspected of being irrelevant, it may be difficult to accumulate a sufficiently large research sample. On the other hand, if one bases selection of the sample on only a few criteria, one runs the danger of failing to include some that are relevant to therapy. For example, only recently has the importance of specifying social class in studies of psychotherapy been appreciated (20, 37).

Since characterization of a population in actuarial terms, however complete, seems insufficient, other modes of description must be considered. Of these perhaps the most obvious is clinical diagnosis. Unfortunately, clinical diagnoses are based on rather vague and overlapping criteria, so that a patient's proper diagnostic label is often in doubt. In one study three well-trained psychiatrists observing patients jointly showed agreement as to the patient's major diagnostic category in only 46% of the cases, and in only 20% with respect to the subcategory (3). An additional difficulty with descriptive diagnostic categories is that they are only loosely related to the major concerns of psychotherapy, the pa-

tient's underlying conflicts and his characteristic interpersonal behavior (10). They may even prove to be completely irrelevant for purposes of psychotherapy, although I do not share this view.

The limitations of the conventional diagnostic scheme suggest directions in which to look for more useful diagnostic criteria. One would be in terms of the patients' motivations and conflicts. Unfortunately, characterization of dynamics must be based on inferences, and these differ depending on the researchers' theoretical preconceptions, so such criteria lose in communicability what they gain in relevance.

Another possibility is to borrow a notion from drug studies and select patients in terms of the "target symptoms" which the particular form of psychotherapy hopes to modify (12), regardless of the clinical syndromes in which they occur. For example, one could study patients suffering from depression, anxiety, or visual hallucinations. Since these symptoms obviously can be expressions of various underlying states, however, the idea does not seem very promising in this form.

If the major target symptoms of psychotherapy are considered to be disturbances in the patients' characteristic ways of perceiving others and behaving towards them, however, then this approach may prove to be very fruitful. A sophisticated and promising example is the scheme of interpersonal dimensions of personality devised by Leary and his co-workers (27).

Having selected the research population and described it as best we can, the next problem is how to divide it into experimental and control samples. There are four theoretically possible ways of doing this. Two exist only in fantasy, but should be mentioned for the sake of completeness. The perfect equivalent control group would consist of patients matched individually in all respects with

those receiving psychotherapy. The matched patients would have identical heredity and life experiences up to the beginning of the experiment, an obvious impossibility. The second impossible method would be to match control and experimental populations, patient by patient, on certain variables believed to be significant, such as age, sex, educational level and so on. Since each additional matching variable greatly increases the size of the population which must be screened, this approach is hopelessly impractical in outpatient studies.

The third possibility is to match by stratified sampling. Groups can be matched with respect to the proportions of patients in certain categories without matching individuals. Even this procedure is ordinarily too cumbersome for an outpatient department, because one cannot wait for a sufficient population of patients to accumulate before starting some form of treatment.³ The final resort is to match by random selection; that is, by assigning patients alternately to treatment and control groups. The assumption is that over the long pull all significant variables will be randomly distributed among both groups, so that differences found in therapy and control groups at the end of the experiment may

3. Two potential ways of increasing the size of the population so as to make stratified sampling possible are by extending the project over time, or by drawing on the populations of several clinics as in the Veterans Administration Study (29). Each, however, introduces its own control problems. Extending the study over time rests on the assumption that time of year is not a significant variable; i.e. that patients presenting themselves for treatment at different times of year are essentially alike. This assumption may not be valid, as considered below. If patients are drawn from many clinics, this increases the number of variables on which they must be matched, such as ethnic group, urban or rural, and so on, which tends to counteract the gain sought by increasing the size of the sample.

safely be attributed to the therapeutic procedure rather than to an unequal distribution of patient variables in the two groups. Whether randomization has been achieved can be checked by simple statistical methods applied to measurable attributes of the populations such as actuarial indices and test scores

All the methods of control of person variables described rest on the assumption that if the control and experimental groups are matched on known variables, or if these variables are found to be randomly distributed throughout both groups, then all other variables which might potentially account for the differences found between control and experimental sample would be similarly distributed. This assumption may not be valid, and this has occasionally resulted in scientific tragedy. To take an example from another field, vast amounts of biochemical work on hospitalized schizophrenics have come to grief because certain biochemical differences between the patients and matched nonhospitalized controls were attributed to schizophrenia, whereas they were really due to institutionalization, with its effects on activity, diet and so on. Recently high hopes were aroused by the discovery that the serum of schizophrenics oxidized adrenaline more rapidly than the serum of normal controls. The hopes were dashed when this phenomenon proved to be caused by vitamin C deficiency in the diets of the patients (1). When this variable was controlled the difference between normals and schizophrenics disappeared. In this example, an unsuspected but crucial difference in situational variables invalidated the matching of person variables in control and experimental groups, leading to an erroneous explanation of the difference obtained.

With respect to control of the patient variable in studies of outpatient psychotherapy, it should be pointed out finally, that though complete matching of con-

trol and experimental groups may be practically impossible to achieve, certain differences between the experimental and control groups need not destroy their usefulness for all purposes. In our study, in spite of every effort to assign patients randomly to the three types of therapy, we found that more lower-class patients were placed in group therapy. The sampling bias made it difficult to interpret a finding that more patients dropped out of group than individual treatment, but did not affect other findings, for example that in all forms of treatment those who scored sickest initially improved most.

Another example of a biased control group in our study is that of patients who drop out of treatment within four sessions. These are self-selected and presumably differ systematically from those who remain in treatment, though the nature of the differences is unspecified. Nevertheless, the drop-outs served as a useful control of negative findings. For example, they showed the same drop in discomfort scores after six months as those receiving various forms of treatment over this period of time (9). This finding permits the conclusion that an improvement in discomfort is not a function of duration or type of psychotherapy received, or of differences in the nature of drop-outs and remainers.

In the Rogers and Dymond studies (15), the equivalent control group was unavoidably biased, in that it was selected from volunteers who were paid to participate in a "research on personality." They did not perceive themselves as sick or needing help—an obviously important difference from the clients in the therapy group. Nevertheless they proved useful for certain types of controls.

An ingenious way of circumventing the whole matching problem is to use each patient as his own control. The patient is observed before and after a time interval in which he receives no therapy. He then receives therapy, after

which the same observations are repeated (15). Changes between the first and second readings are compared with those between the second and third; presumably, differences would be due to the effects of therapy. Own-control designs do not escape practical and theoretical problems imposed by withholding therapy, which will be considered under situational variables. On the whole, when they can be managed, they probably represent a better control than use of an equivalent group, except that they do not control changes in patients related to passage of time. For example, it has often been noted in a clinic that intake falls off during vacation periods (28). Whatever the reasons for this, it raises the possibility that the condition of patients may be affected by factors connected with time of year. If an own-control experiment happened to be set up so that the no-therapy period was in the spring and the therapy occurred in the summer, this would leave open the possibility that differences in patient change in the no-therapy and therapy periods were attributable to the season.

Obviously the proper selection of control groups in any study of psychotherapy is difficult. There may be severe practical restrictions on the researcher's freedom to assign patients to therapy and control groups. He must be alert lest unsuspected bias creeps into these assignments, and must search for important overlooked variables in which the control and experimental groups are not adequately matched or randomized. Even inadequate selection methods are better than none, however, as a means of controlling for the patient variables in psychotherapy.

CONTROL OF SITUATIONAL VARIABLES

From the standpoint of situational controls, the first task is to control for the eventuality that changes in patients observed in the course of a particular

form of psychotherapy are not due to intercurrent life experiences or spontaneous fluctuations in the patient's state. If this can be shown, the question still remains as to whether the changes are really attributable to the aspect of therapy which the researcher hypothesizes to be responsible for them. The chief problem of control in this respect appears to be to distinguish the effects of the therapist's personality or attitude, from the effects of his techniques.

Psychotherapy is only one of many influences which may produce changes in patients. For example, many psychiatric conditions seem to fluctuate spontaneously or to be self-limited (39), and a patient is most apt to seek treatment when he is in a trough. The subsequent improvement may be due to the natural course of his condition. Treatment with younger persons may extend over a sufficient span so that processes of growth and maturation may contribute significantly to the changes observed. Improvement due primarily to extra-therapeutic occurrences, such as a change in job or social relationships, may be erroneously attributed to concomitant psychotherapy. The task of untangling the roles of therapy and life changes is further complicated by the fact that psychotherapy may have contributed to the patient's ability to make such changes.

An obvious way of controlling for whether changes in patients are due to therapy or something else is to compare them with changes in an equivalent group of patients who received no therapy. The no-therapy group has just been considered from the standpoint of control of patient variables. We are now concerned with its use to control situational variables. With outpatients this presents formidable difficulties. The major problem is that an adequate no-therapy control would have to last the same length of time as therapy, to allow the same opportunity for the occurrence

of significant extra-therapeutic events or spontaneous changes in both the control and treatment population. Since six months of therapy is usually considered to be the minimal requirement, a control group would have to go without therapy for six months. Patients present themselves to the clinic because they are in distress and want something done. It is hard to reconcile telling them to wait this long with one's professional conscience. But this is the least of the obstacles. Keeping a sizable number of patients without treatment for so long a period may have an adverse effect on the clinic's community relations. A possible source of a no-therapy control are patients on a waiting list for treatment, assuming that the clinic is so inefficient as to have a sufficiently large one. Experience has shown, however, that patients placed on waiting lists are apt to differ systematically from those who are taken promptly into treatment. Patients who seem more in need of help or who arouse the interest of the interviewer for other reasons tend to receive priority for treatment.

Another difficulty is that most patients who are told to go away and come back in six months for re-evaluation will not do so. In order to avoid a monumental attrition rate, the clinic would need to maintain some kind of regular contact with the patients over this period; and any contact may contain therapeutic components (33), so that the no-therapy control would be violated by this procedure. In addition, patients who are told that treatment at the clinic will not be available for some time, if they are in distress will inevitably seek treatment elsewhere, whether it be from a physician, faith healer, or corner druggist. By the same token, patients in psychotherapy will be less likely to seek out other sources of help. Thus, we would not really have a no-treatment control group as against a psychotherapy group

but two groups, each receiving different kinds of treatment. Since the treatment received by the control group involves ingredients which may also be psychotherapeutic, interpretation of differences in the results obtained in the two groups may be very difficult. An additional potential source of error lies in the fact that rejection of a patient for immediate treatment may affect his attitude in such a way as to influence his scores on self-administered tests of change, a matter discussed more fully below.

The basic difficulty with the no-treatment control is that withholding treatment after interviewing a patient is, in a sense, a positive rejection of him. Psychotherapy is one form of interpersonal relationship, refusal of psychotherapy is another; it cannot be regarded as neutral.

For many purposes a more promising control than withholding treatment entirely is to offer the control group a form of psychotherapy differing in an essential ingredient from that received by the experimental group (42). Since patients in both populations, from their own standpoint, would be receiving therapy, this eliminates the problem of systematic differences in attitude towards the clinic and in the tendency to seek outside help. Since both types of treatment would be conducted for the same length of time, occurrence of spontaneous fluctuations or important intercurrent life experiences would be randomized in the two populations, so any obtained differences in results could be safely attributed to differences in the therapy used.

In this type of control, a major problem is to select and define the therapeutic ingredients in which experimental and control groups differ. Choice must be guided by two considerations: the probability that the variables are therapeutically significant and the ability to define them adequately. However important one may suspect a variable to be, its usefulness for research is limited by the

precision with which it can be described. Wittenborn points out that one of the common failings of research in psychotherapy is failure to define the independent variable. This permits the investigator only to say, "how infrequently his result could be ascribed to chance but the reader is uncertain as to precisely what the result can be frequently ascribed." (43, p. 35)

On the other hand, the researcher must not let himself be seduced into selecting aspects of therapeutic technique for study mainly because they can be easily described. Too often such aspects prove to be therapeutically unimportant, and the result is a beautifully designed and reported experiment which fails to disprove the null hypothesis. Precision has been gained at the expense of significance.⁴

In the Hopkins project we gained the impression from pilot studies that one important difference between therapies might be the amount of contact between psychiatrist and patient, so we incorporated this into our research design (19). Group and individual therapy were more nearly equated in amount of treatment contact than either was with minimal treatment. At the time of the first re-evaluation, patients in group and individual therapy had had approximately the same number of sessions, 15.8 and 17.7 respectively, as compared to 9.3 sessions for patients in minimal treatment. Since group therapy and individual therapy differed in at least one major definable way, namely that in the former several patients are present simultaneously, in the latter only one, the design also permitted determination of possible differences in effects of this variable. By our measure of social ineffectiveness, both group patients and individual therapy patients improved significantly more than minimally treated

patients. This would seem to support the hypothesis that amount of treatment contact, whether in a group or individually, significantly affects improvement in social ineffectiveness, although an alternative hypothesis cannot be entirely ruled out, as noted below.

The most important, and unfortunately the least understood, situational variable in psychotherapy is the therapist himself. His personality pervades any technique he may use, and because of the patient's dependence on him for help, he may influence the patient through subtle cues of which he may not be aware. Dr. David Rioch tells an amusing example of a patient of his who was always depressed in the treatment interviews except on five occasions when he seemed quite bright and alert. This puzzled Dr. Rioch until he reviewed his notes and realized that on these five mornings, and on no others, he himself had taken benzedrine.⁵

It is obvious that the therapist and therapy variables cannot be completely separated. It is unlikely that a therapist can conduct different types of treatment with precisely equal skill or that his attitudes towards them will be identical. Therefore, differences in results obtained by two forms of therapy conducted by the same therapist may be due to therapist rather than treatment variables, especially since the faith of a therapist in a form of treatment may account for much of its efficacy (7). In our psychotherapy study the psychiatrists disliked minimal treatment. They gave it reluctantly and felt that they were shortchanging the patients. The patients remained just as long in this type of treatment as in the other two, suggesting that they were not as lacking in confidence in it as the doctors. It is possible, however, that this difference in the doctors' atti-

4. See, for example, (38).

5. Personal communication.

tudes may have contributed to the finding that patients improved less in social ineffectiveness under the minimal treatment conditions. This example illustrates how difficult it is to control adequately for the influence of the therapist in the absence of fuller knowledge about the role of his personal attributes and attitudes in determining the outcome of treatment. Even though minimal treatment was less effective than group or individual treatment in the hands of three different therapists, we cannot be sure that this was not due to differences in their attitudes to the different approaches, though we believe this to be unlikely.

Evidence that the personal qualities of the psychiatrists were not irrelevant to the results of our study, however, is that one of the three showed a tendency, which did not reach acceptable statistical significance, to obtain better results than the other two with all three forms of treatment by both criteria—discomfort and ineffectiveness. Unfortunately, our project was not designed to elucidate aspects of therapists' personalities related to their therapeutic success.

It is clear that the achievement of better ability to identify and control therapist variables warrants a high priority in psychotherapeutic research. Two studies which represent promising beginnings in this regard are that of M. B. Parloff, who showed that of two therapists of roughly similar and equal training, the one who was able to establish better social relationships also established better therapeutic relations (31), and the series of studies of Whitehorn and Betz, who found that psychiatric residents of similar training could be placed in two classes on the basis of their relative degree of success with schizophrenic patients, and that these classes could be distinguished by certain patterns of scores on the Strong Interest Inventory (4).

CONTROL OF RESPONSE VARIABLES

Any attempt to consider control of the response variable in psychotherapy at once threatens to involve one in the tangle of questions as to what is meant by improvement in psychotherapy. For purposes of the present discussion I shall take the position, without attempting to defend it, that the aim of psychotherapy, as one of the healing arts, is to help the patient feel better and function better. The type of functioning which psychotherapy tries to improve is social behavior in its broadest sense; that is, the patient's ability to establish mutually satisfying relationships with others (32).

Before turning to considerations of these criteria, however, it may be well to pause a moment on another set of response variables. These are changes in the patient's behavior in the interview situation, including, for example, certain autonomic responses, the content of his verbalizations (30, 34), and formal aspects of his verbal behavior (35). Studies of changes in these variables as functions of the activities of the interviewer in a single interview circumvent many of the problems of control of situational factors discussed above. They eliminate problems of the role of intercurrent life experience, or outside therapy, or spontaneous long-term fluctuations in the patient's condition. Detailed studies of patients' responses in the interview situation are yielding much valuable information. The relevance of this information to psychotherapy, however, depends on the establishment of its relationship to long-term improvement in the patients' feelings and behavior, which is still far in the future.

Returning now to what I propose to regard as the ultimate criteria of improvement, let us first consider control problems connected with evaluation of the patients' social functioning. Since it is ordinarily impossible to observe the

patient *in situ*, as it were, estimates of his social effectiveness must be inferences based on reports of patients and other informants, though impressions gained from these can be supplemented, confirmed, or called into question by observations of the patient in the treatment situation. Parenthetically, behavior of patients in therapeutic groups may be more useful for this purpose than their behavior in a private interview, since the group is closer to the interpersonal situations of everyday life (40).

The major control problem is how to minimize biases in the reports and in the observer. The former are best controlled by using at least one informant besides the patient. Presumably another informant will not have precisely the same attitude towards the patient and psychotherapy as the patient does. Comparing and contrasting the information from both sources should enable the raters to reach a more accurate evaluation of the actual state of affairs than relying on either alone.

With respect to rater bias, since the ratings are based on interviews, it is not practically possible to conceal from the raters the kind of treatment the patient has had since the previous rating. This could conceivably be done by transcribing the interviews, having one set of persons edit out all clues as to what treatment the patient had received, a second set check to make sure this has been done, and a third set make the ratings of change. In the present primitive state of our knowledge of psychotherapy, the enormous labor required to yield this level of control of rater bias can probably more profitably be expended in other ways. Moreover, it deprives the interviewer of face-to-face contact with the patient, and thus prevents his taking advantage of non-verbal cues, which may seriously handicap him. In our study we tried to guard against rater bias arising from knowledge of pa-

tients' treatment in three ways. The first was simply to keep this possibility always in mind. The second was to base the final global rating of social ineffectiveness on many sub-ratings of the patient's behavior in different situations, including the interview. Our social ineffectiveness scale permitted ratings on a maximum of fifteen types of ineffective behavior and nineteen categories of social situation, including the interview itself. Patients were only rated, of course, on the behavior and situations for which data were available. As a final safeguard, each interview was rated independently by the interviewer and a concealed observer, and the joint rating was arrived at by a conference between the four raters, two having rated the patient and two the other informant.

The pros and cons of ratings arrived at by conference versus those arrived at by arithmetical combination of individual ratings are complex (18). In general, against the conference ratings is urged the danger that one person may unduly influence the total result, so that the conference would merely confirm the opinions of its most powerful member. We checked on this possibility in one of our studies by correlating ratings of individual conferees made before the conference with the conference rating. All correlations were within the same range, indicating that no one member dominated the group's judgment (24).

In favor of the conference method may be offered that it enables each participant to modify his impression in the light of information presented by the others, so that the rating finally reached by the group should be better than that of any individual in it. On the other hand, they may hear more information than they can digest and evaluate, which may impede their ability to make a valid collective judgment (23).

On balance we thought that conference ratings were probably more valid

than those obtained by averaging individual ones. We did insist that each rater rate the patient before coming to the conference to help him withstand the pressures of the other members.

Turning to the other major criterion of improvement, change in the patient's feelings and attitudes, the only way of tapping these is through his reports, direct or indirect. If the patient can clearly perceive the significance of the information he gives, the question arises of controlling for factors influencing his statements other than his internal state. Indirect measures, which more or less conceal from the patient the significance of his responses, shift the control problem to the validity of their interpretation.

In using scales which are relatively transparent to the patient such as a symptom check list or even Q-sorts which yield such measures as self-ideal discrepancy, one must always keep in mind the possibility that the patient is telling the rater what he wants him to hear, so that changes in scores may be due more to changes in his attitude to the observer or to treatment than to genuine improvement. In Rogers and Dymond's study, for example, the own-control clients who were placed on a sixty-day waiting period of no therapy were divided into two groups: the attrition group who stayed for less than six sessions of therapy subsequently and those who stayed for more than six sessions. On all measures the attrition group showed more improvement over the wait period than those who later accepted therapy. This is interpreted to mean that the attrition group showed more tendency to spontaneous recovery (16). Another possibility exists, which is that at the second testing the remainders wanted to show that they still felt the need for treatment; the attrition group, that they did not want further treatment. Thus the remainders would tend to indicate that they still felt sick, and the attri-

tion group that they did not. Needless to say, such distortion need not be deliberate or conscious. This interpretation would be consistent, I believe, with the Rogers and Dymond findings on the two Q-sort measures (self-ideal relationship and adjustment score), which do not fully disguise from the patient what the experimenter is looking for. I am not sure whether it could also account for the similar results on the TAT, but do not believe that even this highly indirect approach entirely excludes this possibility.

We found in our study that at the start of each evaluation period patients who remained in treatment until the next re-evaluation had higher discomfort scores on the average than those who dropped out before the next evaluation. This is similar to the results obtained by Rogers and Dymond, and suggests the same thing, namely that scores on a discomfort scale are partly the patient's means of communicating that he wishes further treatment.

Many indirect measures of the patient's subjective state have been devised in order to circumvent this type of problem. Projective tests such as the Rorschach or TAT permit measurements of changes in the patient's attitudes through the use of communications, the significance of which is hidden from him. Since projective tests yield permanent records, the points at which they were given can easily be concealed from the rater, eliminating bias based on knowledge of whether or not the patient has had treatment. An inescapable limitation of scores on projective tests is that they do not bear an obvious relationship to clinical improvement. Therefore, they must eventually be validated against other measures of patients' subjective state and behavior. For example, in the Rogers and Dymond study, scoring the TAT one way gave results consistent with other measures but scoring it in a different way, though it yielded

another set of relationships to the treatment variables, gave results that bore no relationship to the other measures of improvement (5, 17).

Implicit in the discussion so far is that scores on measure of response variables may be influenced by the conditions under which they are measured, or by the measuring instrument itself. This problem, which exists even in the physical sciences, assumes major proportions in the evaluation of patients' reports of improvement. As already indicated, these may be affected by the patient's attitude towards therapy or the therapist. There remain to be considered the possible effects of the tester's expectancies, of the form of the test, and of its repetition.

With all measures except the strictly objective ones, and possibly even in these, the tester may influence the scores in accord with his expectations through a process which may be analogous to operant conditioning (14).⁶ The cues which mediate this influence may be so subtle as to escape the awareness of both interviewer and subject. Although the limits of the effects of operant conditioning are not known, it certainly exerts more effect than has been generally realized. Examples are the way Freud's patients fabricated infantile memories to conform with his theory of the etiology of neuroses (11), the recent work by Salzinger and Pisoni who showed that within ten minutes it was possible significantly to increase the number of affective statements made by schizophrenic patients (34), and Murray's analyses of therapy protocols which demonstrated rapid shifts in frequency of certain content categories of the patients' productions in accordance with the ther-

apists' values, even when the latter thought he was non-directive (30).

To complicate matters further, the patient may influence his own subjective state by hearing his report of it. If, in response to factors in the test situation, he says he feels better or worse than he "actually" does, his feelings may change to conform with his behavior, as William James observed long ago in a somewhat different context (22, p. 463).

The form of the test may influence the patients' scores. In our study the correlation between patients' global estimates of improvement and their scores on the discomfort scale was only 0.65. Apparently it made a difference whether the measure was itemized and written, or global and oral. An itemized scale perhaps makes the patient more cautious on the one hand and, on the other, reminds him of complaints which had slipped his mind.

The possible effects of mere repetition of the test must also be considered, especially those due to the patients' greater unfamiliarity with the test and test situation on the first occasion than on subsequent ones (21). He may be puzzled by the test itself or made uneasy by other factors of the situation, which can adversely affect his scores. These influences are apt to be less strong the second time he takes the test, yielding "better" scores. We found that there was a marked average drop in discomfort scores the first time the scale was re-administered, and that on the average this drop was maintained over the following two years. The most probable explanation of this phenomenon would be that the first scores on the Discomfort Scale were artificially heightened by the patient's general uneasiness in an unfamiliar situation, so that we were measuring not so much the effect of six months of treatment as the effect of greater familiarity with the test and the situation. It was possible to control for

6. Note a recent and relevant review of the literature: Krasner, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, 55, 148-170.

this by taking a group of patients at the two-year follow-up interval, giving them a placebo to take for two weeks, and then re-administering the Discomfort Scale. We found a drop of the same order of magnitude between this fourth and fifth administration of the scale, following the placebo, as there was between the first and the second, following psychotherapy (13). The drop in response to the placebo obviously could not be explained by increasing familiarity with the implement. This finding has implications for the relationship of psychotherapy to the placebo effect which are irrelevant here. In this context it is cited as an example of controlling for the effects of repetition of a test.

Discussion of control of response variables in psychotherapy would be incomplete without mention of the importance of follow-up studies to determine long-term effects of treatment. Evaluation of any form of treatment is obviously inadequate in the absence of information as to the duration of its effects. The longer the study continues, the greater is the problem posed by attrition of the sample. This is influenced by patients' attitudes towards the treatment received, and by the kind and amount of information sought by the investigator. The importance of the patients' attitudes may be illustrated by the fact that after two years we were able to obtain re-interviews with 90% of the patients who had originally accepted therapy, but on 33% of those who had dropped out of treatment. Presumably the feeling of the former group towards the Clinic was much more favorable.

The effect of the amount of information desired is suggested by the fact that our most strenuous efforts brought back only 56% of the treated patients for personal re-interviews after three years, whereas Saslow (36) reports a return of about 80% after four to six years, to written requests for limited information.

Thus the conductor of a follow-up study is faced with a variety of choices as to how best to expend his resources. He must balance considerations of completeness of sample against the relative value of various types of information obtained in different ways, and so on, but these questions need not concern us here

DISCUSSION

In conclusion I should like briefly to review certain general considerations about controls, which seem to come up repeatedly in clinical research. The first concerns replication of findings. No findings, however striking, are more than tentative until they have been replicated. Replication, incidentally, need not involve actual repetition of the study. If the population is large enough, the same result can be achieved by dividing both experimental and control groups in half, and using one set as a test of the reliability of the findings obtained with the other.

Replication with a fresh sample tests the adequacy of the description of the variables in the original study and the accuracy of the original observations. This is particularly important in a field like psychotherapy where so much is still unknown. Thus it is not surprising, though perhaps a little disconcerting, to find that repetitions of studies at the same clinics, with presumably similar populations and therapists, have failed to reproduce certain findings that possessed high statistical reliability (29, 41). Failure of others to replicate a finding may lead the original researcher to discover that he had failed to make explicit an important experimental condition. Attempted replication by others also helps to establish the extent to which the original finding can be generalized to populations and settings differing in various ways from the original ones.

Perhaps the most important value of replication is that it guards against *ex post facto* reasoning. There is no limit to the ingenuity of the human mind. It seems to be literally impossible to present a person with a set of data that are so random that he will not be able to read a relationship into them.⁷ In psychotherapy if an experiment seems to demonstrate a certain relationship between therapeutic variables and changes in the patient, the experimenter can always make an hypothesis to explain it. This is a necessary and desirable first step to further research. A common error, however, is to offer the observed relationship as proof of the hypothesis. This circular reasoning can be escaped only by making an explicit prediction on the basis of the hypothesis and then seeing if the prediction is borne out with a fresh sample.

While replication is ordinarily highly desirable, this is not to say that every finding should be replicated. Since, especially in research on therapy, replication may involve many months of work, and time and energy are limited, the investigator must ask himself whether the tentative relationship he thinks he has discovered is important enough to justify the effort of replication, or whether his time would not be better spent looking for more significant data. If he reaches the latter decision, is he justified in publishing the unreplicated finding? I believe he is, to make it available to others, as long as he does not present it as more than suggestive.

A general point about controls, which is obvious to statisticians but seems difficult for some clinicians to grasp, is that statistical methods of control can be applied to a relatively small sample. The size of the sample needed to achieve any given level of significance varies directly with the variability of the responses and

the range of characteristics of the patients in the experimental and control groups, and varies inversely with the magnitude of the difference between the groups at the close of the experiment, assuming that they were matched at the beginning.⁸ Also, it is possible validly to generalize findings obtained with a small population to a very large one, as demonstrated by public opinion polls. All that is required is that the small sample be truly representative of the larger one; that is, that important variables show the same relative frequency distributions in the two groups. Of course, the greater the discrepancy in size between the sample and the total populations, the greater the care needed to assure its representativeness.

Statistical measures of significance may be misleading in that a statistically significant finding need not be significant in the non-technical sense of the term. The discovery of a very low correlation between variables which achieves high significance because the groups involved are large, indicates, to be sure, that some relationship is present, but it may be so weak as to contribute practically nothing to an understanding of the phenomena under study. The central question posed by such a finding is whether pursuit of the lead is likely to unearth a relationship of sufficient importance to justify the effort. An analogy might be the discovery of a low grade of ore. The decision as to whether to try to extract the metal from it would depend on the estimate of the work involved and the potential value of the metal. The Curies used tons of pitchblende to obtain a few grains of radium. They would not have made

7. Personal communication from Alex Bavelas.

8. Kramer and Greenhouse (26) have prepared tables indicating how large experimental and control groups must be in order that given amounts of difference between them will achieve given levels of significance.

a similar effort to extract an equal amount of lead.

Failure to demonstrate a significant relationship between two variables does not prove the absence of such a relationship. The statement that a proposition has not been proven to be true is not the same as the statement that it has been proven to be false. Some of my colleagues accuse me of being a therapeutic nihilist when I point out that it has not yet been demonstrated that different forms of therapy lead to significantly different results. They hear this as an assertion that no such differences exist, instead of an attempt to point out an area requiring research.

Many factors may obscure the existence of a genuine relationship. Wittenborn (43) points out that if the sample is not normally distributed, this increases the standard error of estimate, decreasing the possibilities that the differences between experimental and control groups will meet a statistical criterion for significance. This may lead to a real difference being overlooked. He suggests statistical measures for testing and correcting for this error. Significant differences may also be obscured by insensitivities or errors of the measuring instruments. In research on psychotherapy, the therapist, through unfamiliarity with a new technique, may fail to get results, not because the technique is valueless, but because he uses it poorly. "These are the factors our tender-minded, but not therefore unscientific, investigator bears in mind. He stresses that 'not statistically significant'—like the Scotch verdict 'not proven'—permits us to return the hypothesis on trial to the arms of those who love it, rather than at once chopping off its head." (6, p. 57)

Though controls enable the investigator to state the level of confidence of his finding, they do not insure a correct interpretation of it. That the serum of hospitalized schizophrenics oxidizes adrena-

line more rapidly than the serum of non-hospitalized controls was established with a high level of certainty and replicated, but this did not prove that the difference was due to schizophrenia. It turned out to be due to differences in the diet of the two populations. We can assert that the chances are better than 95 in 100 that patients will show more improvement in social effectiveness after six months of weekly individual therapy than after six months of minimal treatment, but the interpretation of this difference remains open. It would take another study to determine if the difference is best explained by the fact that patients in minimal treatment had less therapeutic contacts, or by some other factor, for example that minimal treatment was devalued by therapists and patients. These examples illustrate that often one cannot control for a variable until one thinks of it. The automatic use of controls is no substitute for thought.

Since controls are always relative, and since energy expended in establishing controls is diverted from that devoted to seeking new insights, their proper use is not solely a question of applying the correct methods, but involves exercise of judgment concerning what level of control should be attempted. How much control to strive for is determined by the state of development of the subject and the potential importance of the finding to be checked. Sometimes it may be better to accept a poorly controlled tentative finding as a source of leads for potentially more penetrating or definitive studies, than to divert time and energy to trying to increase its level of confidence. Efforts to use a level of control not warranted by the state of the problem may be as hampering to good research as failure to use controls that are possible.

Preoccupation with controls, moreover, is apt to guide the selection of questions for study, not by their signifi-

cance, but by the ease with which they can be investigated. One is reminded of the familiar story of the drunkard who lost his keys in a dark alley but looked for them under the lamp post because the light was better.

What is most needed in research on psychotherapy is originality of thought and courage to grapple with important issues, setting up as much control as is feasible. Each experiment should lead to another which is an improvement over its predecessor. In this sense a bad experiment is better than none, and several are better than one. Unless one makes the original crude experiments, no progress is possible.

REFERENCES

1. Angel, C., Leach, B. E., Martens, S., Cohen, M., & Heath, R. Serum oxidation tests in schizophrenic and normal subjects. *Arch. Neurol. & Psychiat.*, 1957, 78, 500-504.
2. Appel, K. E., Lhamon, W. T., Myers, J. M., & Harvey, W. A. Long term psychotherapy. In *Psychiatric treatment*. Baltimore: Williams & Wilkins, 1953. Pp. 21-34.
3. Ash, P. The reliability of psychiatric diagnoses. *J. abnorm. soc. Psychol.*, 1949, 44, 272-276.
4. Betz, Barbara J., & Whitehorn, J. C. The relationship of the therapist to the outcome of therapy in schizophrenia. In N. S. Kline (Ed.), *Psychiatric Research Reports 5*. Washington: Amer. Psychiat. Assoc., 1956. Pp. 89-105.
5. Dymond, Rosalind F. Adjustment changes over therapy from thematic apperception test ratings. In C. R. Rogers and Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univ. Chicago Press, 1954. Pp. 109-120.
6. Edwards, A. L., & Cronbach, L. J. Experimental design for research in psychotherapy. *J. clin. Psychol.*, 1952, 8, 51-59.
7. Frank, J. D. The dynamics of the psychotherapeutic relationship, 1. determinants and effects of the therapist's influence. *Psychiat.*, in press.
8. Frank, J. D., Gliedman, L. H., Imber, S. D., Nash, E. H., & Stone, A. R. Why patients leave psychotherapy. *Arch. Neurol. & Psychiat.*, 1957, 77, 283-299.
9. Frank, J. D., Gliedman, L. H., Imber, S. D., Stone, A. R., & Nash, E. H. Patients' expectancies and relearning as factors determining improvement in psychotherapy. *Amer. J. Psychiat.*, in press.
10. Frank, J. D., Margolin, J., Nash, Helen T., Stone, A. R., Varon, Edith, & Ascher, E. Two behavior patterns in therapeutic groups and their apparent motivation. *Human Relat.*, 1952, 5, 289-317.
11. Freud, S. *A general introduction to psychoanalysis*. New York: Horace Liveright, 1920.
12. Freyhan, F. A. Psychomotility and parkinsonism in treatment with neuroleptic drugs. *Arch. Neurol. & Psychiat.*, 1957, 78, 465-472.
13. Gliedman, L. H., Nash, E. H., Imber, S. D., Stone, A. R., & Frank, J. D. The reduction of symptoms by pharmacologically inert substances and by short term psychotherapy. *Arch. Neurol. & Psychiat.*, 1958, 79, 345-351.
14. Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, 68, 409-416.
15. Grummon, D. L. Design, procedures, and subjects for the first block. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univ. Chicago Press, 1954. Pp. 35-52.
16. Grummon, D. L. Personality changes as a function of time in persons motivated for therapy. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univ. Chicago Press, 1954. Pp. 238-255.
17. Grummon, D. L., & John, Eve S. Changes over client-centered therapy, evaluated on psychoanalytically based thematic apperception test scales. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univ. Chicago Press, 1954. Pp. 121-144.
18. Hamburg, D. A., Sabshin, M. A., Board, F. A., Grinker, R. R., Korchin, S. J., Basowitz, H., Heath, H., & Persky, H. Classification and rating of emotional experiences. *Arch. Neurol. & Psychiat.*, 1958, 79, 415-426.

19. Imber, S. D., Frank, J. D., Nash, E. H., Stone, A. R., & Gliedman, L. H. Improvement and amount of therapeutic contact: an alternative to the use of no-treatment controls in psychotherapy. *J. consult. Psychol.*, 1957, 21, 309-315.
20. Imber, S. D., Nash, E. H., & Stone, A. R. Social class and duration of psychotherapy. *J. clin. Psychol.*, 1955, 11, 281-284.
21. Jacobs, A., & Leventer, S. Response to personality inventories with situational stress. *J. abnorm. soc. Psychol.*, 1955, 51, 449-451.
22. James, W. *The principles of psychology*, Vol. 2. New York: Holt, 1890.
23. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univ. Michigan Press, 1951.
24. Kelman, H. C., & Parloff, M. B. Interrelations among three criteria of improvement in group therapy: comfort, effectiveness, and self-awareness. *J. abnorm. soc. Psychol.*, 1957, 54, 281-288.
25. Kline, N. S., Tenney, A. M., Nicolaou, G. T., & Malzberg, B. The selection of psychiatric patients for research. *Amer. J. Psychiat.*, 1953, 110, 179-185.
26. Kramer, M., & Greenhouse, S. W. Determination of sample size and selection of cases. *Proceedings of the conference on the evaluation of pharmacotherapy*, Washington, D. C., September 19-22, 1956, in press.
27. Leary, T. *Interpersonal Diagnosis of Personality*. New York: Ronald Press, 1957.
28. Lhamon, W. Time and rhythm in psychosomatic relationships. In P. Hoch and J. Zubin (Eds.), *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953. Pp. 244-255.
29. Lorr, M. Progress and problems in research on psychotherapy. Paper read at VA-Univ. Conf., Univ. Maryland, November 14, 1957.
30. Murray, E. J. A content-analysis method for studying psychotherapy. *Psychol. Monogr.*, 1956, 70, No. 13 (Whole No. 420).
31. Parloff, M. B. Some factors affecting the quality of therapeutic relationships. *J. abnorm. soc. Psychol.*, 1956, 52, 5-10.
32. Parloff, M. B., Kelman, H. C., & Frank, J. D. Comfort, effectiveness, and self-awareness as criteria of improvement in psychotherapy. *Amer. J. Psychiat.*, 1954, 111, 343-351.
33. Rosenthal, D., & Frank, J. D. Psychotherapy and the placebo effect. *Psychol. Bull.*, 1956, 53, 294-302.
34. Salzinger, K., & Pisoni, Stephanie. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, 57, 84-90.
35. Saslow, G., Matarazzo, J. D., Phillips, Jeanne S., & Matarazzo, Ruth G. Test-retest stability of interaction patterns during interviews conducted one week apart. *J. abnorm. soc. Psychol.*, 1957, 54, 295-302.
36. Saslow, G., & Peters, Ann D. A follow-up study of "untreated" patients with various behavior disorders. *Psychiat. Quart.*, 1956, 30, 283-302.
37. Schaffer, L., & Myers, J. K. Psychotherapy and social stratification: an empirical study of practice in a psychiatric outpatient clinic. *Psychiat.*, 1954, 17, 83-93.
38. Semon, R. G., & Goldstein, N. The effectiveness of group psychotherapy with chronic schizophrenic patients and an evaluation of different therapeutic methods. *J. consult. Psychol.*, 1957, 21, 317-322.
39. Shepherd, M., & Gruenberg, E. M. The age for neuroses. *Milbank Memorial Fund Quart.*, 1957, 35, 258-265.
40. Stone, A. R., Parloff, M. B., & Frank, J. D. The use of "diagnostic" groups in a group therapy program. *Int. J. Group Psychother.*, 1954, 4, 274-284.
41. Sullivan, P. L., Miller, Christine, & Smelser, W. Factors in length of stay and progress in psychotherapy. *J. consult. Psychol.*, 1958, 22, 1-9.
42. Watterson, D. J. Problems in evaluation of psychotherapy. *Bull. Menninger Clin.*, 1954, 18, 232-241.
43. Wittenborn, J. R. Critique of small sample statistical methods in clinical psychology. *J. clin. Psychol.*, 1952, 8, 34-37.

The Research Strategy and Tactics of the Psychotherapy Research Project of The Menninger Foundation and the Problem of Controls¹

LEWIS L. ROBBINS, M.D., AND ROBERT S. WALLERSTEIN, M.D.²

In our initial publication we stated that "The purpose of the Psychotherapy Research Project of The Menninger Foundation is to study the process and course of psychotherapy in order to increase our understanding of how psychotherapy contributes to changes in patients suffering from mental illness." (5) The questions and the methods of this study of the determinants of change in psychotherapy are based on psychoanalytic theory, a number of the fundamental tenets of which constitute the assumptions from which the hypotheses of this project are derived. These assumptions were stated in the previous article as follows:

1. Mental illness derives from otherwise insoluble intrapsychic conflicts.
2. These conflicts are in large part unconscious.
3. Intrapsychic conflicts are related to early childhood experiences and represent inadequately resolved infantile conflicts.
4. Prior to the onset of clinical illness the intrapsychic conflicts are handled

1. The manner of conceptualization of the strategy and tactics of our project in this presentation owes much to the suggestions and refinements offered by our consultant, Wayne H. Holtzman, Ph.D., whose help is gratefully acknowledged. In addition, we gratefully acknowledge the critical and helpful interest of our consultant, John D. Benjamin, M.D., whose influence affects our operations and our conceptualization of them in this as in every other aspect of our project's work. We also express our appreciation of the generous support of the Foundations' Fund for Research in Psychiatry and The Ford Foundation.

2. Co-chairmen, Psychotherapy Research Project of The Menninger Foundation.

through the idiosyncratic patterning of impulse-defense configurations, character traits and perhaps more or less ego-syntonic symptoms which together make up the personality structure of the individual.

5. Through varying combinations of inner and outer stresses (sometimes clearly discernible as 'precipitating events') the previously utilized methods of maintaining homeostatic equilibrium fail and symptoms or ego-dystonic character traits, or both, appear.
6. The patterning of the symptoms and associated ego-dystonic character traits reveals important elements of the inner conflicts, the ways the ego tries to cope with them as well as important aspects of the fundamental character organization of the individual. (5)

Within this framework, we have set out, not to "validate" psychoanalytic theory but to refine and test a series of hypotheses derived from within that theory, bearing on the problems of treatment process and treatment outcome. These hypotheses would be those that *implicitly* guide our regular clinical treatment planning and treatment practice. The research job would be to make these clinical prognostications and the inferential process on which they are based as *explicit* as possible and to translate these judgments into verifiable predictions, with their assumptive base, the contingencies to which they are linked, and the evidence necessary for their confirmation or refutation all clearly stated.

Our primary model for this is the evaluation process (psychiatric case summary, psychological testing, and social history) from which treatment recommendations and treatment planning reg-

ularly derive in our clinical setting. As stated in our previous publication (9), at the evaluation conference at which a clinical consensus is sought,

certain clinical judgments such as statements of prognosis with or without the recommended treatment, and predictions regarded as contingent upon events which may be expected to transpire within the therapy or in the life of the patient are usually made explicit. . . . In addition, the evaluation process also results in a series of implicit judgments—the selection of the most suitable type of treatment, the decision about hospitalization, even the diagnosis itself. All these are statements of predictive and prognostic import as well as statements referring to current status and current needs.

The assumptions upon which these predictions are based are for the most part tacit and only to a certain, though somewhat variable, degree directly stated in each case under consideration. At conferences, evaluative statements are offered, buttressed by a statement of some of the outstanding factors that, in the particular case, seem to contribute to the formation of that judgment. Sometimes feeling, intuition, or experience is the expressed basis.

We have attempted to define *systematically* and *explicitly* these prognostic and evaluative criteria, whether explicitly or only inferentially used. This conscious focusing and elaboration of these factors can thereby be used to throw more light on the elements that determine change (and govern prediction) in psychiatric treatment and on the validity of the assumptions that underlie these judgments.

In thus basing our research endeavor on making more explicit, and subjecting to scientific scrutiny the assumptions and hypotheses governing our regular clinical practice, we have committed ourselves to a naturalistic approach, to the study of ongoing clinical practice and the theory that guides it, uninfluenced, so far as possible, by the research study itself. In accord with this principle we undertake no manipulation of treatment methods for research purposes nor do we set up untreated control groups. Although our design allows for "experiments" in the study of contrasting variables or contrasting phenomena, these are experi-

ments created by the *selection* of clinical instances sought in order to fill certain specifications, rather than by the *manipulation* of the clinical material. In the usual sense of the word, this is then not an experimental project. Rather it is an intensive clinical study of a number of patients being treated by various forms of psychoanalytically based psychotherapy, studied by research methods designed not to alter the therapeutic process in any particular.

Our principles of case selection stem from these strictures. The cases to be studied are chosen from among those which are treated at The Menninger Foundation in accordance with our regular clinical practice. Treatment is prescribed and carried out by the staff independently of the research study and in no way is there any alteration of the patient's treatment because we have elected to study his particular case. In fact, neither the therapist nor his supervisor knows that the case being treated is one of those selected for our research until the termination of the treatment.³ The major requirement in selection of the case is that individual psychotherapy of some type be the principal modality of treatment employed, whether the patient is an outpatient, is hospitalized or is in the day hospital. This excludes those seriously disturbed patients for whom hospital treatment encompassing a variety of modalities represents a major part of the total therapeutic effort, even though they may concomitantly be treated in individual psychotherapy. Our cases have also been selected to provide an even distribution between males and females and are all between the ages of 19 and 45.

3. On occasion the supervisor, and more rarely, the therapist, has cause to find that a patient of his is a "research case" while the treatment is in progress. This is a lapse in project structure that we strive constantly to guard against.

In addition, fifty percent of the patients in our series are being treated with psychoanalysis, a considerably higher percentage than is found in the ordinary distribution of treatment modalities in our practice. The decision to stratify the sample by this higher proportion of cases from this modality was based on two grounds. It happens that due to clinical considerations in our setting (the existence of a psychoanalytic institute and the need for many psychoanalytic cases as "control cases"), among cases treated by psychoanalysis there are a much larger number for whom detailed—daily—process notes are available. Also, because of the inherent nature of psychoanalytic treatment, more information is elicited in depth about intrapsychic configurations than is usually obtained in face-to-face psychotherapies. Thus we will have a larger number of cases on whom the more extensive and penetrating data regarding intrapsychic processes is obtained. The decision that half our cases should be in psychoanalysis and half in face-to-face psychotherapy does not express any intent to consider these as comparison groups. We are not "comparing" these psychotherapeutic modes with one another but rather trying to study the applicability, the efficacy, and the nature of each. Our total sample for the first round of our research is 42 cases.

We conceive of this first phase of our Psychotherapy Research Project as an outcome-process study at an *exploratory* stage. In forthcoming publications (7, 10) we state our conviction about the conceptual inseparability of outcome and process considerations, although we recognize that operationally these twin aims are often opposed and decisions of relative emphasis upon one or the other must be made by each research project at each stage of its progress. In those articles we state too our rationale for the choice of methods that we think ought,

in the present stage of our work, to be geared to a broad exploratory study of both outcome and process aspects of psychotherapy. Certainly our ultimate objective is the one we share with most workers in the field of psychotherapy research—better knowledge of the processes by which changes occur as a result of treatment. The major research goal with our initial series of cases is the development of precise hypotheses which will be more specifically inter-related to the variables found most significant for the process of change in psychotherapy. This present phase of our study we consider therefore to be primarily one of hypothesis-finding and hypothesis-refining. It will be the task of subsequent phases to set up the more precise and detailed inquiry with fewer patients studied at more frequent intervals that will lead to hypothesis verification or refutation. Equally important to this beginning phase of our work is the forging and testing of instruments to be used in the many different tasks involved. To this end a number of procedures and forms by which data gathered can be ordered and treated in a variety of ways have been devised. These will be described below.

There are many levels of scientific inquiry at which efforts to describe therapeutic processes may be made. The first, the level of ordinary observation, either about an apparently singular patient, or aspect of a patient, which is not yet brought to the point of established abstraction or generalization to other cases, is that of the usual clinical case report. It is not the specialized concern of organized project research such as is under consideration here. The second level involving ordering and systematization of observations according to a theoretical framework within which these are conceptualized (thereby allowing application and generalization to other cases) is the level at which most clinical psychother-

apy research projects are conducted. It is the chief level at which our own project currently operates but with upward projections and advance provision for multi-level studies. Within this level, our qualitative descriptions are made not only of observable behavior and manifest aspects of the patient's organization and functioning but more importantly include judgments and inferences regarding the intrapsychic processes which the manifest behavior reflects. These judgments and inferences are not confined only to assessments of shifting constellations within the patient but also encompass what occurs in the dyadic relationship of patient and therapist, as well as the psychological impact of reality events in the patient's current life situation upon him.

The judgments and inferences on the defined variables are made within the frame of reference of the basic assumptions (our "givens") outlined at the beginning of this paper and the postulates and hypotheses derived from them. One group of these postulates has to do with those factors in the *patient*, in his personality organization, and in the structure and dynamics of his illness that bear on his psychotherapeutic course and outcome. A second group of postulates has to do with those factors in the *treatment* and in the *therapist* deemed relevant to the course and outcome of the psychotherapy. These are concerned with the effect on the therapeutic course of the specific technical maneuvers of the psychotherapy, as well as the effect of the basic attributes and attitudes of the therapist, and of the climate within which the patient-therapist interaction takes place. A third group of postulates is concerned with those factors in the *life situation* of the patient which, on the one hand, influence the treatment course and outcome, and on the other hand may themselves be influenced by the patient, thus reflecting the treatment course and outcome.

The third level of clinical study, that of attempted quantification and formalization of those aspects of the essentially qualitative clinical data which lend themselves to mathematical concepts, also finds representation in our project, mainly in the form of ordinal scaling from which patient profiles are derived, to be used in selecting subjects for factorial experiments. Quantitative methods are of course of circumscribed usefulness for clinical investigation for they are necessarily subject to at least two types of distortion. Transposing rich and complex clinical data into measurable categories inevitably means lifting these data out of context and oversimplifying them, thus distorting them. Such distortion is inherent then to some degree in our arrangement of multi-dimensional psychological phenomena into a linear sequence for purposes of ordinal ranking. Secondly, when clinical data are fitted to mathematical models there is additional distortion resulting from the fact that the assumptions of the model are never exactly met by the facts of nature. Statistical manipulations of clinical data can never accurately yield all that the model itself implies. The amount of over-all distortion depends on how closely the model selected corresponds to the data to which it is applied, as well as how well the data can be represented by the categories without undue simplification. Each research project of course makes its own decisions as to the amounts and kinds of distortion that can be tolerated without unduly violating the model or jeopardizing the usefulness of the clinical conclusions that can be drawn.

The fourth method of clinical study, the experimental, involving the assignment of patients to varieties of psychotherapy in accord with research, rather than purely clinical criteria; involving deliberate manipulative devices within the treatment situation; and at times even

TABLE I
HIERARCHY OF METHODS IN CLINICAL RESEARCH

<i>Level</i>	<i>Purpose</i>	<i>Problem</i>	<i>Data</i>	<i>Variables</i>	<i>Method</i>
I.	Hypothesis finding Isolating of variables	Ordering phenomena Q.—How can these phenomena be un- derstood?	Unique case Special phenomena	To be defined	Description Classification Conceptualization
II.	Hypothesis finding and formulating	Discoverer relation- ships Q.—Are A & B re- lated?	Manifestations se- lected for observa- tion Coded data Judgments	Categories Processes inferred and conceptualized	Description Serial observations Crude quantifications & correlations Ordinal scaling
III.	Hypothesis refine- ment Prediction testing	Identify relevant fac- tors in relationships; determining condi- tions Q.—Under what con- ditions does <i>r</i> (re- lationship) hold? What new phenom- ena does it imply?	Scores, measured char- acteristics	Isolatable dependent & independent vari- ables	Statistical studies Factorial designs Tests of predicted re- lationship
IV.	Verification Extension of hypotheses	Identification of invar- iant determinants Q.—Given A, does B follow?	Phenomena produced under controlled conditions	As above	Laboratory studies Predictions verified or refuted

involving the withholding of therapeutic effort, falls outside the scope of naturalistic study in our project, as we have defined and limited it.

In a previous publication (6), a member of our group (Sargent) defined the position of our project within a hierarchy of research methods in psychotherapy research. She designated four levels, *clinical*, *process (or co-variant)*, *factorial*, and *experimental* distinguished by the state of knowledge and the precision with which hypotheses can be expressed (which in turn determines the methods which are appropriate). These were then depicted in the following tabular form which differs only in small particulars from the conceptual framework advanced today.

In terms of this table, Sargent stated that "the Psychotherapy Research Project, as we envisage it, is centered largely in Level II, but rests solidly upon clinical research at Level I, and is equipped to test certain hypotheses at Level III."

This then is the conceptual framework. The essential data of this project consist of the ordering of our clinical facts and inferences around the three groups of coordinate variables which reflect the three major sets of postulates of our project, and which we designate as (1) patient variables, (2) treatment variables and (3) situational variables. These variables have been chosen because they are each felt to be relevant to the course and outcome of psychotherapy within a psychoanalytic frame of reference and each is assessed at three major points in time in terms of the treatment of each case. The interrelationships discerned among these variables as they co-vary in time, will become the specific "hypotheses" of our research, the goal of our exploratory, hypothesis-finding round, and the subject matter of a subsequent definitive hypothesis-testing round.

The patient variables are those factors in the patient, in his personality organi-

zation and in the structure and dynamics of his illness which we believe relevant to the course and outcome of his treatment. Some of these such as the nature and severity of his symptoms or the intensity and manner of manifestation of anxiety are directly reportable by the patient. Others such as the nature of the core neurotic conflict, or the patterning of ego defenses, or the anxiety tolerance are complex assessments arrived at via the clinical inferential process from a consideration of the total available clinical data in each case. These patient variables that we have selected for study have been ordered and described in our previous publication (9) under the following list of headings:

Form B: Assessment of Relevant Patient Variables

- I. Sex and Age
- II. Anxiety and Symptoms
 1. Anxiety
 2. Symptoms
 3. Somatization
 4. Depression and guilt feelings
 - a. Depression
 - b. Conscious guilt
 - c. Unconscious guilt
 5. Alloplasticity
- III. Nature of Conflicts
 1. Core neurotic problem
 2. Current life problem
- IV. Ego Factors (and Defenses)
 1. Self-concept
 2. Patterning of defenses
 3. Anxiety tolerance
 4. Insight
 5. Externalization
 6. Ego strength
- V. Capacities Factors
 1. Intelligence
 2. Psychological-mindedness
 3. Constitutionally endowed aspects of ego strength
 4. Capacity for sublimations
- VI. Motivational Factors
 1. Honesty
 2. Fee
 3. Extent of desired change
 4. Secondary gain

- VII. Relationship Factors
 1. Quality of interpersonal relationships
 2. Transference paradigms
- VIII. Reality Factors
 1. Presence of "neurotic life circumstances"
 2. Adequacy of finances to the treatment requirements
 3. Attitudes of significant relatives
 4. Physical health

The treatment variables consist of those factors in the treatment and in the therapist deemed relevant to the course and outcome of the psychotherapy. Like the patient variables, the treatment variables comprise a number of kinds of assessments. They include the technical interventions that characterize the formal elements and the process of the psychotherapy; they include as well those attributes and attitudes in the professional personality of the therapist and those aspects of the climate within which the patient-therapist interactions take place, that are felt to significantly influence the therapeutic course and outcome. These treatment variables are ordered and described in a forthcoming publication (3) under the following list of headings:

Form T: Assessment of Relevant Treatment Variables

I. Formal Elements of the Treatment

A. Treatment Modality

1. Variations determined by the nature of the patient's illness
2. Variations determined by reality events
3. Variations determined by the therapist
4. Variations determined by the supervisor

B. Treatment Instructions, General and Special

C. Special External Circumstances that Affect the Treatment Course

D. The Basic Techniques of the Treatment

1. Suggestion
2. Abreaction
3. Manipulation
4. Clarification
5. Interpretation

E. Types of Subject Matter

1. The patient's past life and memories
2. Reconstructions and memories of the conflicts of infancy and childhood.
3. The patient's current life situation
4. The events of the treatment hour
5. Dream and fantasy material
6. Style and form of the patient's expression
7. Accompanying affect of the patient's expression

F. The Goals of the Treatment

1. The goals of the patient
2. The goals of the therapist
3. Changes in goals

II. The Process of Treatment

A. Life of the Patient During Treatment

B. Significant New History

C. Course of Treatment

1. Major themes
2. Major transference paradigms
3. Changes in symptoms
4. Changes in impulse-defense configurations
5. Changes in manifest behavior patterns
6. Structural changes in the ego
7. The acquisition of insight, and its relation to changes in attitudes towards the self, others, things, and the illness
8. Style and form of communication

D. Termination of Treatment

III. Variables in the Therapist and the Climate of the Patient-Therapist Interaction

A. Qualities of the Therapist

1. Skill and experience
2. Need for supervision and how he uses it
3. Personal style and attitudes
4. Countertransferences
5. Age and sex

B. Climate of the Therapy

The third set of variables consists of those significant situational and social-interactional factors which both influence and are influenced by the patient and his treatment. Since each situational factor

must be assessed with respect to its individual psychological impact on the patient and on his treatment, the relevant situational variables are ordered and described in a forthcoming publication (8) in two-dimensional form under the following list of headings:

Form S: Assessment of Relevant Situational Variables

- I. Situational Factors
 - A. Background
 1. Cultural
 2. Religion, ethics and values
 3. Education
 - B. Interpersonal Relationships
 1. Family relationships (parental)
 2. Family relationships (own family)
 3. Friends, coworkers, colleagues, etc.
 - C. Marital-Sexual
 - D. Living Situation
 1. Home arrangements
 2. Financial
 3. Responsibilities
 - E. Occupation
 1. Stability
 2. Occupational demands
 3. Compensation
 - F. Community and Leisure Time
 1. Recreation and avocation
 2. Group participation
 3. Civic and cultural activities
 - G. Physical
 1. Appearance and physique
 2. Handicaps
 3. Somatic illness
- II. Each of these situational factors is evaluated in respect to the following *psycho-situational dimensions*:
 - A. Degree of relevance in the individual life
 - B. Amount of stress
 - C. Degree of support
 - D. Degree in which the situation is a crucial conflict area
 - E. Extent of opportunity for self-realization, growth, autonomy
 - F. Congruence of the situation to needs, interests, capacities
 - G. Degree of situational mutability

The basic data from which the assessments of these three groups of variables are derived consist first of the clinical material which is gathered as part of our regular evaluation process on the basis of which treatment is prescribed and planned. At this initial point in time the research utilizes only routine clinical material from the clinical record. The patient is studied next at the time of termination of the treatment and finally a follow-up study is made approximately two years after the termination of treatment. These termination and follow-up studies are special research additions to the routine contacts with the patient and with those involved in his life and his treatment.

Form A consists of all this *direct data* gathered at each of these vantage points of observation. In the initial study it is the clinical assessment of the patient prior to entering formal psychotherapy. It consists of our entire usual evaluation procedure and includes the psychiatric case study, psychological testing and a summary of social work contact with the significant family members.

For the termination study the following data are added: copies of the monthly *psycho-social* progress notes; copies of sample interviews from the daily process notes; the psychotherapy discharge summary which summarizes the course of the psychotherapy and the patient's current clinical status; terminal psychological testing of the patient; and narrative accounts of the research interviews held at the time of treatment termination with the therapist, the therapy supervisor, the patient, and with those individuals significant in the patient's life during treatment who may include hospital personnel, foster care family, or the patient's own family.

For the follow-up studies, patients are asked to return to Topeka for a complete reassessment which includes extended psychiatric interviewing of the patient,

psychological re-testing and social work interviewing of the significant relatives as comparable to the initial clinical evaluation as the altered circumstances allow (this being a follow-up of a person no longer a patient and no longer seeking help).

This material gathered as Form A is at the level closest to the primary data. In Form A historical material, observable behavior and clinical inferences are recorded in the way in which these are ordinarily combined and subsequently condensed and abstracted in clinical records. This is the first of three levels of data organization and classification in our project.

The next level is that of qualitative generalization from the Form A material. The first step on this level consists of ordering the Form A data by clinical judgments in accordance with the subcategories of the three sets of relevant variables, patient, treatment and situational. From a consideration of all the available Form A information, at each point in time, a pair of clinician-judges formulates and agrees on a *clinical* assessment of each variable under consideration. This is formulated as a clinical judgment and written out essay style as a descriptive statement. In so far as possible both the specific data in Form A that have led to that judgment and the theoretical assumptions within which those data have been organized and given meaning are specified. Three different pairs of researchers make these judgments, a different group for each set of relevant variables. These judgments are then spelled out in the three forms outlining these groups of variables already listed.

Form B, Assessment of Relevant *Patient* Variables, consists of the inferences drawn from the clinical data of Form A concerning the series of patient variables which define and determine treatment choice and prospect. At each of the

three points in time (initial, termination and follow-up) a separate pair of researchers independently reviews the clinical data for its portion of the study and assesses these variables for each patient.

Form T, Assessment of Relevant *Treatment* Variables, is a similar detailed statement about the factors in the treatment and in the therapist which interact with the patient variables to determine treatment course and outcome. Unlike the patient variables which can be equally well assessed at each point of time, the treatment variables are first viewed at the time of termination although recommendations and predictions about them are made as part of the initial study of the patient. Further information about them gained at the time of follow-up generally either accents or modifies the major view obtained at termination.

Form S, Assessment of Relevant *Situational* Variables, is the counterpart of Forms B and T, conceptualizing those situational and social-interactional variables which represent an assessment of the patient's interaction with his environment prior to, during and subsequent to treatment. Form S is a tripartite form consisting of (1) a descriptive series of judgments about the psychological significance of the patient's life circumstances, (2) a tabular schematization of this qualitative data, and (3) an objective list of events and changed circumstances during and after treatment. It differs from the other forms in that it is a two-dimensional form which attempts to differentiate the objective situation with respect to its individual psychological impact upon the patient. Emphasis here is upon the impact of the situation on the patient and equally upon the patient's role in determining the situation. The information utilized in assessing the situational variables is derived from the Form A material at all three points of time.

Following completion of the assessment of patient variables at the time of the Initial Study, Form C (the individual Prediction Study) is prepared. This is likewise a clinical document prepared by the same two clinician-judges who have made the Form B patient variable assessments. Form C consists of a series of opinions, predictions, and prognostic statements regarding treatment course and outcome, together with the supporting reasons for them, all based on the assessment of the patient variables, and their grouping and juxtaposition in accord with the theoretical frame of reference within which they have been defined. Form C statements are built around the five aspects of treatment which have been selected as being crucial to its course and outcome. These are (1) the variations in the transference, (2) the changes in symptoms, (3) the changes in general behavior patterns, (4) structural changes in the ego, and (5) the acquisition of insight, and its relation to changes in attitudes toward the self, others, things, and the illness. Form C states in detail the treatment recommendations, the assumptions about the patient and about the nature of the treatment on which the predictions are based, and predicts the expected vicissitudes of the patient's responses during the course of therapy and the anticipated outcome of the therapy. At termination and follow-up the changes in the five areas covered in Form C are compared with the predictive statements about them made at the time of the Initial Study. Thus in Form C the patient variables which have been assessed in Form B are manipulated in accordance with our theoretical framework and our specific postulates in order to make predictions about treatment course and outcome. Form C, like Form B, consists then of a series of clinical judgments made by a pair of clinician-judges based on available clinical information. What distinguishes these

forms as research documents from ordinary clinical operation is that the categories deemed relevant are pre-selected and defined and in each instance a specific and explicit clinical judgment is entered in each category, together with a statement about the historical and observational data that have led to that judgment and the theoretical assumptions within which those data have been organized in order to lead to that judgment.

Following the completion of Form C, Form P (the Formalization of the *Predictions*) is prepared. This is a recasting of the clinical predictive statements into a series of formal predictions; each such statement contains the assumptions on which it is based, the contingencies if any to which it relates, and a predicted outcome so stated as to be readily confirmed or refuted. In doing this, individual predictions are of course removed from their full clinical context which, as already indicated, produces a degree of distortion of the clinical meaning. However, this distortion is kept to a minimum by the hierarchical ordering of the predictions in a way which indicates their interrelatedness, by stating the contingencies which apply to them, by estimations of their probability, and by level of confidence statements. An example from one of our cases of such a formalized prediction, stated in an "if-then" form, is the following: "*If* the psychoanalytic treatment unfolds as anticipated, *then* a major transference problem will be between his overly compliant and obedient way of relating himself on the one hand, and his extreme fear of free association (with the implication of absence of controls on his thinking) on the other hand."

All of these Form P formalized predictions are then arranged in categories determined two ways, (1) by the nature of the limiting conditions (contingencies and assumptions) to which they relate and (2) by the nature and areas of the

predicted events and judgments to which they point. It will then be possible from the individual prediction study of each patient to ascertain those kinds of predictions and those areas of prediction that can be forecast relatively accurately from those in which this is not possible within the limits of our existing theoretical constructs and available clinical knowledge. More important, since at each step along the way the observations and assumptions on which the particular judgments are based are explicitly stated, it becomes possible to reverse the inferential process in regard to any prediction in order to examine how and why it went astray (or for that matter, how and why it seemed to come out). From just such considerations, our concepts of what the relevant variables are in regard to psychotherapeutic course and outcome will be strengthened, modified, or refuted as the case may be. We hope to be able then to delineate more accurately which the really crucial variables and theoretical assumptions prove to be (in the sense of making for more accurate prediction); which variables overlap or are even aspects of one another; and which prove to be unimportant.

The means for gathering this evidence necessary to validate or invalidate each specific prediction is embodied in the next Form, Form E (Evidence), a form in which commitment is made in advance to the evidence—factual or judgmental—which will be accepted as confirming or refuting each prediction. Form E accordingly is prepared and executed following the completion of Form P. Each prediction of Form P is recast into a number of true-false statements covering a range of choices in respect to future events (i.e., the patient will get married, the patient will have a stable work situation) or judgments (i.e., the patient will be less guilt-ridden, the patient will have greater anxiety tolerance). Each prediction of Form P is made to yield a series

of two or three plausible statements of differing possible outcomes. In addition to these individual Form E statements, uniquely devised for each patient, there is a series of some 70 core Form E statements, the same for every patient. These were derived from our first few cases as statements about kinds and areas of prediction that seemed to come up repetitively and that one ought to be able to make some statement about in every case. Examples of such core statements are (1) the patient now shows a greater capacity to control impulses or (2) alloplastic behavior occurred to a degree that neither psychotherapeutic nor management interventions could adequately cope with it. Having these identical statements in a core form used with all patients in addition to the individual statements creates a group where direct interpatient comparison is facilitated, and where outcomes in regard to specific predictions can be related to differences in kinds of patients, kinds of treatments, or kinds of situations.

All of these Form E statements in each case (individualized as well as core statements) are then keyed by the initial study group, the clinician-judges who made the original assessments of relevant patient variables and clinical predictive statements in that case. These judges key each statement in accordance with their original predictions, marking each statement "true," "partly true," or "false." They thereby state in advance of the treatment the evidence to be obtained upon termination and follow-up that will confirm or refute each of the predictions made in clinical context in Form C, and abstracted and formalized in Form P, thus avoiding the easy assumptions of *post hoc* reasoning that can retrospectively explain almost any outcome. Another copy of the same Form E is filled out subsequently by those who study the case upon treatment-termination and at the time of follow-up. The answers to

each of these statements (both about events that have transpired and judgments that can be made) that are given by the termination and follow-up groups can then be directly compared to the predictive statement about how they thought it would come out, made by the initial study group.

All of the preceding manipulations of data have been within the level of qualitative description. Some of our data lend themselves to some extent to the next level of abstraction or generalization, that of quantitative description. An example of this is our use of ordinal scaling derived from a modified application of Fechner's method of paired comparisons worked out by Sargent. In our previous publication (4) this modification was described as follows:

Fechner's interest was in psychophysical parallelism; the methods he devised were for the purpose of testing relationships between subjective sensory thresholds and the absolute values of the physical world . . . In psychotherapy research the problem is different; there are no external absolutes against which to check. We are, however, concerned as Fechner was with the scaling of subjective judgment, and for this purpose one of the methods he bequeathed is useful. The clinician is not accustomed to "measuring" a patient; but it is not an unfamiliar operation to judge a patient as "more than" or "less than" another patient on such variables as "insight" or "anxiety."

As stated in that article, this modified method of paired comparisons can be applied to judgments on any dimension pertinent to patients, treatments, or situations which may at any stage of the research seem worth investigating as long as the *degree* as well as *presence* of the factor is regarded as clinically significant. Thus a variable like "anxiety tolerance" is suitable to such a method of study, whereas "self-concept" which we only describe qualitatively is not. Even when a variable is suitable there is nonetheless the distortion resulting from arranging multi-dimensional psychological phenom-

ena into a one-dimensional sequence (that of degree alone), in this fashion. Nevertheless, these forced choices, resulting from the pairing of every patient with each other in batches of manageable size, along whatever dimensions are amenable to this method, yield ordinal scales and useful profiles that can be derived from them, which will serve for the study of similar and of contrasting characteristics through factorial designs.

Thus the original clinical data have passed through a series of studies in which generalizations are made in respect to processes within individual patients, beginning with details of history, and behavior and going to more and more central dynamic attributes. In addition generalizations are made through the group of patients by which we attempt to find common principles characteristic of kinds of patients, kinds of treatments, and kinds of situations, and to investigate their interactions.

Upon the completion of the gathering of all the data at the three points in time and the completion of the various forms for each of the three sets of variables, the qualitative and quantitative analysis of the data collected in respect to our basic postulates and hypotheses will then be undertaken. As yet we have no results to share with you. As already stated, our total sample for this stage of our research is set at 42 cases. Each patient is followed through the course of a psychotherapy averaging three years in duration, with a follow-up on an average of two years after termination. Thus far thirty-nine initial studies have been completed. Eleven of these patients have come to termination and in three cases we already have follow-up data. We have thus been in a position to have already put all our methods and instruments to trial. Whether they will yield what we think they promise us we will be able to say in a future report. We hope from our observation of the broad

sweep of the psychotherapeutic course from our strategic vantage points to generate and subject to preliminary test fruitful hypotheses about the specific interrelationships of the many variables in the patient, in his treatment, and in his life situation that do prove relevant to psychotherapeutic course and outcome. These specific crucial hypotheses will then be the basis of further and more detailed study of the mechanism of change in psychotherapy, using methods and observation posts appropriate to such a more detailed and specific inquiry.

Implicit in our original commitment to a naturalistic rather than an experimental approach to the study of psychotherapy is the absence from our research design of the kinds of manipulations inherent in setting up "normal" controls, matched groups of treated cases and non-treated cases, etc. But this choice of approach does not abrogate in any way our responsibility to tackle the ever-thorny problem of controls in psychotherapy research. It forces us rather to consider afresh (1) what to control and (2) how to control it; by what specific control methods.

Our over-all premise has been to base our major efforts towards control more on appropriate *selection* of clinical material than on its *manipulation*. In other words, we align ourselves with Cronbach's "correlation psychologists" though sharing his distaste for the current dichotomy between "experimental" and "correlational" methods. (2) Thus our strategy has been to set up criteria for the selection of instances which represent the hypotheses we wish to test and to look for the "experiments in nature" that provide cases representing the patterning of variables we wish to examine and compare. Linked to this is the deliberate omission of non-treated controls or of normal controls in the usual sense. Non-treated controls are usually set up on a matched basis, in which *custom* dictates

the control of such variables as age, sex, economic status, marital status, duration of illness, formal diagnosis, etc. Since we feel that these simple differential criteria do not distinguish individuals from one another, along dimensions crucial to their psychotherapeutic course and outcome, we bypass these criteria and by selection concentrate our attention on our assessments in depth of the three groups of variables, patient, treatment, and situational, that we do deem *relevant* to the course and outcome of treatment. How we use our assessments of these variables to match patients for similarity or contrast in regard to any grouping of them we will discuss below under the heading, interpatient controls. We do not use these differential criteria to set up non-treated controls since that would mean withholding treatment that is felt to be clinically indicated (at least temporarily) and would therefore violate our decision to study naturally occurring treatment processes in regular clinical practice, without any research alteration of treatment planning or execution. Nor do we set up a so-called normal control group. Often this means simply "not in a hospital" or "not in treatment" which is not a dimension particularly relevant to an understanding of either personality function or the nature of illness.

Still another premise governing the use of controls in our project has to do with the level at which one introduces the concept of control. In experimental work controls are most often introduced at the level of data collection in the form of specifications and restrictions as to what the observer should look for. We choose rather to allow the clinician-observer to work as he is accustomed—to use whatever conceptual categories are natural and congenial to him, and to gather and utilize in his formulations and judgments all the clinical data available to him at each of the points of study. Post-audit controls are then intro-

duced at the level of data classification and analysis, as formalized in our various instruments, in which data are recorded in defined categories. Thus our choice, as in the case of the projective techniques, is to introduce controls at the level of data analysis, and avoid imposing any limitation on the original clinical observations.

Within this conceptual framework we have then addressed ourselves to the "how to control it" question and have selected a number of specific methods of control. The first method, *inpatient control*, that is, using the patient as his own control, is accomplished by the device of the individual prediction study. In many current psychotherapy research projects the concept of the patient as his own control is used in the sense of comparing the amount and kind of change, if any, during a pre-treatment period of time with the amount and kind that then occurs during the period of treatment that is studied. In order to meet the requirements of this design, the patients are usually put on a waiting list for a predetermined period of time before starting treatment and are then restudied at the end of this waiting period, just prior to the actual onset of the psychotherapy. The arguments against considering this pre-treatment period as a valid reflection of the course of the illness without treatment to be compared then with its course during treatment have been cogently stated by others (including Dr. Frank in his presentation today).

We use the concept of the patient as his own control in a different sense, in the sense propounded by Gordon Allport in his exposition of ideographic vs. nomothetic research inquiry (1). We thus set out to make a longitudinal individual predictive study in each member of our research sample. At the time of the initial study, specific and detailed predictions are made regarding treatment

course and outcome, based on the clinical assessment of the patient variables considered within our theoretical conceptual framework. We have discussed this in the description of the way the clinical judgments of Form B (Assessment of Relevant Patient Variables) and of Form C (The Individual Prediction Study) are made, and how these clinical statements are transformed into verifiable predictions in Form P, each prediction having its assumptive base, and the contingencies to which it relates stated as clearly as possible. As already stated these predictions are made both in respect to subsequent events and subsequent judgments. At the same time the evidence that will be necessary to confirm or refute each prediction is predetermined and specified. This evidence is then sought at the time of the termination and follow-up studies through Form E, as already described.

Having specified for each prediction the assumptions on which it is based, and the contingencies to which it relates, we are in a position to uncover the bases for the confirmation or refutation of the prediction. We can learn to what extent and in what areas both the underlying assumptions and the specific variables chosen do or do not permit accurate prediction, thereby narrowing the areas for future, more specific (and more definitive) investigation. By setting down the predictions, the necessary evidence, and the assumptive base all in advance, we have sought to introduce control into our observations and to avoid the temptations of *post hoc* reconstruction and rationalization.

The second control method that we use, that of *interpatient control*, is based on the assessments via Forms B, T and S of the three groups of relevant variables, patient, treatment, and situational, in each case. As already stated some of these variables lend themselves to comparative evaluation by clinician-judges in

our modified application of the method of paired comparisons. The resulting ordinal rankings are converted into profiles for each patient which facilitate interpatient comparison and contrast, and permit the selection of patients who are alike in respect to certain of our variables and dissimilar in respect to other variables. Thus some variables can be controlled while the variability of others is investigated. It will be possible to select patients representing extremes of rank in regard to specific variables or combinations of variables that seem relevant to specific hypotheses we may wish to test. Such profiles, useful though they are for these purposes, are of course limited by the fact that not all the variables we postulate to be relevant to psychotherapeutic process and outcome can be included in them.

Our third control method is the parallel and independent study of psychological tests. A battery of psychological tests is employed at each of the three stages of the study (Initial, Termination and Follow-up) to assess the patient variables and, to the extent possible, the treatment variables. From the assessment of the patient variables from the test data, individual prediction studies are made in each case, exactly comparable to the prediction studies made from the clinical psychiatric material. At the subsequent points in time, the nature and extent of change in the variables under study is again judged, from the tests alone, parallel to, but independent of, the concomitant clinical restudy.

Such psychological test studies have the advantage of not requiring any knowledge of preceding clinical or of preceding test data. The comparative judgments of the change in status of the variables studied can be made after they are independently assessed at each point in time. In this way, cross-sectional judgments can be made independently of knowledge of earlier stages. Therefore,

we can have assessments which are not only independent of clinical data, but also completely independent of previous assessments using the same test methods.⁴ It is not as possible for each successive clinical study to be as uninformed in respect to the data which precede it. Those doing the Termination and Follow-up studies do not know what predictions have been made, but it is not possible in the course of gathering clinical information at Termination and Follow-up to avoid knowledge of how the patient viewed himself and how others viewed him in the course of time.

Thus, through the prediction study aspect of our design, the use of the patient as his own control is employed in both the clinical and the psychological test studies. The two studies being conducted concurrently but independently act both as checks on themselves and on one another. These two control methods, along with the method of paired comparisons, are built into the design of the entire project, and are employed in relation to each of the three groups of variables, patient, treatment and situational, that are studied over the course of time.

In addition, from time to time we encounter some "inadvertent controls" which are studied whenever available and constitute a fourth method of control. For example, at times patients are seen who, because of finances or geography, are unable to follow the treatment of choice, and for whom an alternate treatment plan is worked out in accordance with their reality situation. In other instances there may be a difference of opinion between the clinical staff which

4. This is a statement of our ideally "complete" design both for the study of the psychological test data per se and for their use as a "control" on the concomitant clinical psychiatric study. Exigencies of time and personnel may force some limitation on the step-wise completeness of these test studies.

had originally seen the patient and prescribed the treatment and the research group in respect to what the treatment of choice should be. In either of these instances, differing sets of predictions (Forms C, P and E) are made out for the treatment actually undertaken, and for any other form of treatment that is deemed ideally more appropriate. For instance, it may be felt that a certain symptom or characteristic of a patient can change only if the patient is psychoanalyzed. If the patient is not psychoanalyzed but treated by some other psychotherapeutic modality, and the change that was postulated to be dependent on psychoanalysis nevertheless does take place, then the bases on which the particular prediction was made would be open to serious question. At the present time about fifteen percent of our cases constitute one or the other type of inadvertent control, thus lending themselves to such study.

The four methods of control just described are built into this first phase of our research. We anticipate that in a subsequent phase of our study we will be able to establish much more precise control criteria in terms of the more sharply defined central variables that emerge from our exploratory round. Patients can then be very specifically selected in accordance with their similarity to or their difference from certain profiles or specific groupings of variables. In addition, it might be possible to cooperate with other clinical groups where treatment, perhaps on a basis other than psychoanalytic theory, is employed, and parallel studies in respect to treatment course and outcome can be made, employing similar variables, similar criteria, and similar instruments.

The manifold problems inherent in research into the process of psychotherapy have thus forced us not only to develop new instruments for our tasks but also to develop a design which de-

parts in many ways from the experimental tradition which has advanced scientific knowledge so tremendously in the last century. Nearly all fields of science have been enriched by the experimental method in which the use of the laboratory and its associated methods of control by manipulation have become traditional. Great stress has been placed, too, on the necessity for absolute objectivity on the part of the observer and to this end measuring instruments, photographic plates and many other mechanical devices have been employed. Although we fully subscribe to the concept of "objectivity" in science, the concept as applied to psychology has been narrowly, it might even be said ritualistically applied. In carrying out our undertaking, we are trying to keep open-minded and to be on the alert and ready to try out a "new look."

Certainly man's knowledge in many areas has been advanced greatly even though the laboratory and its usual control methods could not be employed. This has been particularly true in such sciences as astronomy, geology and archaeology. With the discovery of the unconscious, the study of man's psychological processes has become more complicated than ever and our awareness of the tremendous number of variables operative both within the individual himself and between the individual and his environment has made us often feel that the task of studying the determinants of change in psychotherapy is an overwhelming one—and the task of controlling such study well-nigh hopeless.

However, in the past few years courage to undertake this task has been derived from a number of sources. The development of statistical techniques appropriate to the complexities of psychological research and of its control problems has been a tremendous addition to our knowledge and our confidence. Also in recent years we have all come to know

physicists and astronomers somewhat more intimately. In contrast to what we had previously believed, namely that their sciences were most exact, dealt very little with speculation and were uninfluenced by subjective observation, we have discovered that they have to struggle with virtually the identical methodological problems that we do. It should be possible therefore to take courage from this and to feel free not only to employ all that which is useful from the techniques of the past but also to strike out in totally new directions with the confidence that in time if not we, then our successors, will achieve the understanding in this field which we all so urgently seek.

REFERENCES

1. Allport, G. *Personality: A Psychological Interpretation*. New York: Henry Holt & Co., 1937.
2. Cronbach, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, 12, 671-683.
3. Luborsky, L., Fabian. Michalina, Hall, B. H., Ticho, E. & Ticho, Gertrude R., The Psychotherapy Research Project of The Menninger Foundation, Second Report. II. Treatment Variables. *Bull. Menninger Clin.*, 1958, 22, 126-147.
4. Luborsky, L. & Sargent, Helen D. The Psychotherapy Research Project of The Menninger Foundation. V. Sample Use of Method. *Bull. Menninger Clin.*, 1956, 20, 263-276.
5. Robbins, L. L. and Wallerstein, R. S. The Psychotherapy Research Project of The Menninger Foundation. I. Orientation. *Bull. Menninger Clin.*, 1956, 20, 223-225.
6. Sargent, Helen D. The Psychotherapy Research Project of The Menninger Foundation. II. Rationale. *Bull. Menninger Clin.*, 1956, 20, 226-233.
7. Sargent, Helen D. Intrapsychic change: Methodological problems in psychotherapy research. (A condensed version of this paper was read at the Symposium on "Theoretical and Methodological Problems of Psychotherapy Research." American Psychological Association, Sept., 1957.)
8. Sargent, Helen D., Modlin, H. C., Faris, Mildred T., and Voth, H. The Psychotherapy Research Project of The Menninger Foundation, Second Report. III. Situational Variables. *Bull. Menninger Clin.*, 1958, 22, 148-166.
9. Wallerstein, R. S. and Robbins, L. L. The Psychotherapy Research Project of The Menninger Foundation. IV. Concepts. *Bull. Menninger Clin.*, 1956, 20, 239-262.
10. Wallerstein, R. S. and Robbins, L. L. The Psychotherapy Research Project of The Menninger Foundation, Second Report. I. Further Notes on Design and Concepts. *Bull. Menninger Clin.*, 1958, 22, 117-125.

Discussion of Papers by Frank, and Robbins & Wallerstein

JOHN M. BUTLER, PH.D.

The problem of control is so closely allied to problems of experimental design and of experimental design to the problem of the purposes of the investigator that I approach this discussion with some trepidation. I feel it would be presumptuous of me to talk about the design of experiments in any general sense for I take it for granted that all of us here have been educated in a scientific tradition. I would like, however, to comment on what I perceive as significant differences between the two papers just presented, and then discuss in particular some of the implications thereof.

In the first place, Dr. Frank's preoccupation is, I believe it is fair to say, hypothesis testing. For him the purpose of control is to exclude as many alternatives as possible. For Drs. Robbins and Wallerstein, control, in the sense of manipulation of the objects of study for the purpose of excluding alternative hypotheses is clearly subordinate to the purpose of understanding how psychoanalytic psychotherapy or psychotherapy based on psychoanalytic theory contributes to changes in patients suffering from mental illness. Hence, in studying on-going clinical practice they are committed to a naturalistic as contrasted with an experimental approach. It is clear, then that one of the major differences resides in the commitment, in the programs described, to an experimental versus a "naturalistic approach." Nevertheless, a measure of control is clearly presented in both studies although Dr. Frank's is closer to the criterion of being scientific when science is equated with experimental method.

Secondly, there is a distinct difference in the goals of the investigations: Dr.

Frank is interested in improvement in the studies he cites whereas Drs. Robbins and Wallerstein are interested primarily in studying and explicitly describing intra-psychic processes in the interest of developing more precise hypotheses.

It seems to me that from these differences in philosophy and purpose, major differences in my opinion, the difference in methods of attack upon the problems selected by the investigator derives. Let us consider first the research of the investigators at the Phipps Clinic. At once we note that the patients with organic brain disease, anti-social character disorders, overt psychosis, and mental deficiency were excluded from the study. Furthermore, the patients were individually characterized by diagnostic categories, by an inventory covering aspects of their behavior and by initial scores used on scales to measure change.

Now I take it that the exclusions and characterizations represent control measures as well as what are formally called independent variables; that is, they are to vary as the experimental conditions vary; if not, not.

I also take it that the restrictions imposed upon the therapists: considerable experience in individual therapy, one group per therapist under supervision, and equal lengths of trainings at Hopkins, also present control measures used to exclude hypotheses alternative to those that the investigators wished to test.

Finally, we come to the specifications of the experimental conditions:

1. The therapist is to behave in such a way that a relationship is established with the patient which will help him to identify and correct current distortions in his interpersonal perceptions and be-

havior. In helping with this the therapist, in general, stresses:

- a. in individual therapy, the present more than the past.
- b. in group therapy interactions within the group rather than outside events.
- c. in "minimal therapy"—not described except to say that the focus is on the patient's complaints and how he might deal with them.

In addition, I might say that the scales are the actual dependent variables but they collectively represent two constructs called comfort-discomfort and social ineffectiveness-social effectiveness.

Additional control measures were: each therapist gave all three forms of treatment and patients were assigned at random, an effort was made to keep therapist and patient in contact for at least six months.

Let us consider now the structure of this experiment. First, the experimental conditions, i.e., the behavior of the therapist in therapy is described in extremely general terms. We know only what the therapist is to stress in therapy, more than that, we don't know what he is going to do and we don't know what he did when he did do it. In what sense, then, is the study replicatable?¹ Next the question arises: what is the relation of the dependent variables, the scales, to the experimental conditions? The scales measure decrease in discomfort and decrease in social ineffectiveness: what have distortions in interpersonal perceptions and behavior to do with discomfort and ineffectiveness? I ask this question in the sense of asking for a specified set of relations. Why, for example, if the experimental conditions stress interpersonal perceptions and behavior, do not the scales measure interpersonal percep-

tions and behavior? Now I am sure I know the practical answer to these questions but I am asking formal questions. I am fairly sure that the formal answers are first that the study is not replicatable, strictly speaking, and second, that a derived relation between specification of the experimental conditions and the measuring scales, cannot, strictly speaking, be given.

It follows that, in a formal sense, this study does not meet the requirements of experimental method. Alternatives to the hypotheses under consideration cannot be excluded.

What is the basic reason why the study described does not meet experimental canons? Dr. Frank implies the answer when he says that the degree of control is relative to the state of development of a field of study. What this means, in practice, is that all possible alternative hypotheses can never be excluded in the design of an experiment, when they can be, I doubt that experiments are necessary. What is at issue is the basis on which we control. In my opinion, controls, that is to say, experimental regard for alternative hypotheses, should be based upon what is known scientifically; not upon any set of alternative hypotheses which the fertile man of mind can derive. As Dr. Frank rightly remarks we perceive order even in random data and to create order is to create hypotheses. Since any one of us can spawn hypotheses as fish spawn eggs, we necessarily choose, in a rather arbitrary fashion, sets of alternative hypotheses to control unless we have a sound principle for selecting control variables out of the practically infinite number of variables which may be selected. Obviously the selecting principle should be scientific; this means to me simply that the more an alternative hypothesis has a firm empirical or scientific base the more claim it has to be controlled.

1. The conference pointed out that any description of a study makes it replicatable. It appears then, that a material rather than a formal question was being asked here.

The controls used at Hopkins represent, no doubt, good guesses but they are not firmly based on what is well-established scientifically speaking. This is shown by the ingenuity which alternative hypotheses were advanced. I believe that very few of the alternative hypotheses which it was desired to exclude was on a firm empirical or scientific base. It is to be expected that alternative hypotheses always come to mind when a study is not fully controlled or when controls do not have a firm scientific base. I found myself devising what seemed to me reasonable alternative hypotheses as I read Dr. Frank's paper; some were based on imperfectly controlled studies I know of or have done and some are based on my personal experience as a therapist. In short, alternatives were based on fairly uncontrolled experience with no firm ground in experience as the phrase is defined by science.

If you take my discussion to mean I place no credence in the results reported you will be in error. I think it is significant that group and individual therapy resulted in more improvement than "minimal" therapy. And that patients in general had more sessions in group and individual therapy than in "minimal" therapy. If the results had been in favor of "minimal" therapy I would have been quite disconcerted; in fact, I would not have believed them because they would have contradicted what I think I know and because the experiment is not tight enough.

In my view, Dr. Frank has shown quite clearly the difficulties and problems of conducting controlled experimentation in psychotherapy. If you share my view and his that the degree of experimental control possible depends upon the state of knowledge in a field of study, then no doubt it is as obvious to you as it is to me that our difficulties have an intrinsic basis. No amount of sophistication and ingenuity in arranging experiments is in

itself enough to overcome our ignorance. Only intellectual hard labor and the arduous and concomitant task of getting more and more contact with the phenomena of psychotherapy can do that.

Now for the paper of Drs. Robbins and Wallerstein. Like all exploratory studies, the study described is complex and its very complexity to my mind is likely to produce as much or more in the way of hypotheses as of established results. All I can do not being a psychoanalyst is to say that a naturalistic approach to investigation is one toward which I am sympathetic. Their program of study as described gives evidence of considerable thoughtful planning. I would like, however, to make a few comments about naturalistic observation. Traditionally, the naturalist observes rather than experiments, and he observes *in situ* so to speak although this tradition is being changed by the ethologists such as Lorenz and Tinbergen. It is not that the naturalist does not have ideas nor is it so that he is too inept or ignorant to design experiments; rather, he says in effect, "I have ideas but they are not firmly based in an empirically tested theory. Neither are they based firmly on raw fact. They are more or less personal or intuitive so I will just go out and look around to see if my ideas make sense and perhaps to develop some new ideas. Then perhaps I will be ready to evolve the ideas into a theory; if not I will probably have a better idea of what to control in the experiments that I may do."

Let me continue my flight of fancy by imagining that my naturalist has become a psychotherapist who is interested in research. He has had considerable experience and has a more or less systematic set of notions by virtue of having done considerable therapy. We will assume further that he has taken no notes and now decides to become more systematic. He writes down as completely

as he can all that he thinks about therapy and decides to use his writing as a guide for observing the behavior of clients in psychotherapy. This guide may be called his classification system. Like any accurate naturalist he has a means of recording the behavior of his clients. He records what took place when so that when therapy is finished he has a complete record of what happened. That is, therapy in its behavioral manifestations can be reconstructed from his record *in terms of his classification system* or descriptive scheme. Now he looks at a case noticing the relations between his data. He notices what happened within each interview, what did not happen that he expected to happen, etc. He also notices the relations between interviews and the relations between the various classes of behavior over therapy without regard to interviews. Finally, when he has exhausted the relations as far as he can, or when his patience gives out, he performs some complex mental acts of integration thereby developing some ideas, I prefer to call them hypotheses, such as repression, infantile sexuality, transference, social interest, self-concept, self-ideal relationship, denial, and the like. I think I have described, rather crudely to be sure, but I think accurately, the observational basis on which a naturalist clinician arrives at his conclusions, conclusions which have the character of hypotheses. There is one exception, however, My naturalist recorded all of his observations instead of trusting to his memory.

This method, which I think is basically the method of Adler, Freud, Jung, Meyer, Rank, Rogers, Sullivan, and company has obvious disadvantages. There is not control. There are no specifications of conditions. The therapist does his work and notices the reactions of the client. He may introduce specification of "experimental conditions" by describing briefly or extensively what he, as therapist did, or better, taking sound

recordings, or even better, taking movies. However, historically, therapists have done relatively little of this. But the more the better. I might add that I think it better to conceive of the "conditions of the experiment" apart from control measures which embody alternative hypotheses.

If our clinician has described his own behavior in much the same way as he has that of his client, then he is describing what is analogous to the "habitat" of the naturalist rather completely. Once this habitat is described the clinician is in a rather better position than his predecessor. But there is one important lack: a lack which our clinician shares with other naturalists. He has a set of data contained, in this case, on data sheets as well as in his own head. But he has no means of ordering this data except in his own head. This may have some advantages but it seems to me that what is needed is an intervening step which would be extremely useful. That step is an unambiguous method of ordering the data in a manner which makes evident all or most of their interrelations. If this step could be taken, if this method could be developed, then in principle at least, the problem of the ambiguity of inference could be pushed back one more step in the process of drawing inferences from naturalistic observation. From personal, subjective assessment of relations in the data, which in our case are simply description of behavior of client (and perhaps of therapist) we could move back the personal, subjective evaluation to considering the possible meanings of the found relations. And this, I think is one point where just such creative mental acts belong in the process of naturalistic observations and inference. The other point is before the psychotherapy begins where our naturalist pours his creativity into deciding just what seems to be important about psychotherapy and what it is important to observe. I do not think

that such personal creative acts should go into teasing out the relations among the data.

Now I have been cautioned not to talk about my own work so at this point I can only say that on the basis of regarding naturalists in the way just described, it is possible to generate not only one but several methods of ascertaining the relations among the data in a relatively unambiguous fashion leaving only the development and creation of new hypotheses on the basis of clearly defined relations to the creativity of the investigator. Perhaps I can give you a hint by saying that under certain circumstances even the elegant methods of optimal scaling so well described by Bock (1) may be applied to data which faithfully represent the actual behavior of persons in psychotherapy.

Even though the subjective assessment of data relations be made objective, the type of naturalistic analysis I have described is certainly inferior to experimentation when a real choice is possible. For example, our naturalist may have clearly exhibited order in his data. He then draws inferences and generates hypotheses or even systems of thought on the basis of the order. But, I, or you, on being shown the same order may generate different hypotheses or systems of thought. Clearly, alternatives are not excluded.

At this point I have come full circle. In my view experimentation in our field cannot really exclude alternative hypotheses because in reality control depends

upon positive knowledge; that is, upon previous scientific work. At present, controls are pretty much on "you lays your bets and you pays your money" basis. Naturalistic observation and analysis certainly leads to non-unique hypotheses and conclusions even with the "improvement" I have alluded to. My conclusion is that at the present stage of development of our field there is no real choice as between naturalistic, clinical approaches to investigation, however, true it may be that experimental research is our ultimate goal and criterion. It seems to me that our best approach to the conger of problems embodied in the word control is, to paraphrase a distinguished physical scientist, to do our damnest with our minds in our area of investigation without worrying too much about formal problems of about how "scientific" we are.

REFERENCES

1. Bock, R. D. *The Selection of Judges for Preference Testing*. Psychometrika, 1956, 21, 349-366.
2. Butler, J. M. *The Analysis of Successive Sets of Data*. 1955. Vol. I. No. 1. Counseling Center Discussion Papers. The Univ. of Chicago Library.
3. Butler, J. M. *Measuring the Effectiveness of Counseling and Psychotherapy*. Personnel Guid. J., 1953, 32, 88-92.
4. Butler, J. M. *On the Structure of Groups and Institutions*. 1956. Vol. II. No. 13. Counseling Center Discussion Papers. The Univ. of Chicago Library.
5. Teale, E. W. *The Fascinating Insect World of J. Henri Fabre*. New York: Fawcett World Library, 1956.

Problems of Controls

EDITORS' NOTE: The presentation of the discussion here and in the subsequent three topic areas is an edited version of the complete discussion periods. In the original plan for the publication of the proceedings it had been arranged that a summary of these discussion periods would be provided, rather than the publishing of the verbatim transcript itself. The latter, in conferences of this nature, is often much too discursive to be of interest or value even to the participants themselves, let alone the reader who has not had the opportunity of hearing this in the original context. However, the participants at this conference verbalized many important points in the sequence of the discussion, and it was felt that a summary would not do justice either to the content or the process of the discussion periods. It was decided, therefore, to provide an abridged version of the actual discussion by eliminating some of the less central comments and trimming down some of the duplication of ideas presented. Also, all the procedural comments, such as the chairman's remarks, have been eliminated.

As a result of this editing, the reader will probably notice some disjointedness in presentation and there may be a few references to previous comments which have been eliminated. On the other hand, we believe the presentation in its present form provides a more vivid and useful written account of the actual discussion than would have been possible in summary form. There was a most vital and stimulating interaction in the original discussion which we hope has not been lost in this edited version.

DR. FRANK: I am very grateful for Dr. Butler's comments, many of which I will have to think about a great deal more, because they were quite profound.

I agree completely that this study does not meet the requirements for an ideal formal experiment. I believe it can be replicated. It all depends on what you are trying to replicate. The hypothesis

that could be replicated is that the amount of treatment contact may be positively related to the amount of improvement in patients' social effectiveness, regardless of what the therapist does. This part it seems to me could be replicated.

DR. WATKINS: Actually, Dr. Butler has given us a great deal to think about, and mainly to push us to further clarify what we ourselves are driving at.

The main point I would like to make is to advance a somewhat different point of view about the value systems of science. Dr. Butler said two things. First, he equated what is scientific and what is experimental; and then he made the related point that naturalistic observation is okay where you are limited to it, but it is naturally inferior to experimental observations. I guess our position is that we don't think that is necessarily so. Both naturalistic and experimental methods have their appropriate areas of applicability. They may not necessarily be in a position to replace each other and we don't look at one as necessarily a way station on the way to the other. We don't necessarily see that the end point of all scientific work is experiment with manipulation and with control of variables. We have backing from other fields in saying this, not only the ethological work that he called attention to of Lorenz and Tinbergen, but there are obviously many other scientists who also don't work experimentally. Astronomy is a classical example, as is geology, as are any number of sciences which depend predominantly on the capacity to observe, and to be able to put the observations to-

gether in some kind of meaningful framework. What we are saying, then, is that as in geology and astronomy, there are experiments in nature and our question is how do we go about studying them, and what is our end point if it is not an experimental one? Our key word, again, like the astronomer, is prediction. Do we have phenomena out of which we can predict, and out of which we can predict in ways that we hope will control for us the dangers of bias or subjectivity or however you want to call them that seemingly bedevil the problem of prediction? The intervening step that Dr. Butler called for, we think we have in the concept of making specific predictions about a psychotherapeutic course and outcome. Further, in regard to each prediction we require a statement of the contingencies it is based upon, so that we have the statement of what has to be present first in order for this prediction to have meaning, that is, so that we know what is necessary in order for us to either refute it or establish it. Then by stating the assumptions on which the prediction is based, both the assumptions concerning events and the theoretical assumptions derived from our conceptual frame of reference, we are in a position always to retrace this step. So upon finding that a prediction has gone awry or gone well, we can trace the process by which we arrived at it and maybe we can rectify our concept of the relevance of the particular variables to the kinds of predictions one can make from them. More than that, we intend to specify in advance what evidence it will be necessary to have at the end in order to establish or refute the prediction. This, as we said in the paper, is to avoid the problem of retrospective reconstruction in accord with the way we predicted the particular outcome would come about.

I would like to make one other general point. We are not predicting only future *events*. That is, if such and such

happens in the course of therapy, then this event will be seen or that event, but we are predicting as well to *judgments* that will be made at that future time. We are committed to a position that not all judgments can necessarily be traced back to behavioral manifestations. In the end, we are left again with the judgment of clinicians, with all the support it can find in all the behavioral phenomena that lend it support, ordered together within a conceptual framework and a theoretical position that you have which too is your clinical judgment. How good that is depends on how often other people like yourself paying attention to the same data, can come up with similar judgments and what predictions can be made from those judgments which can become subject to the same kind of refutation or confirmation.

DR. BUTLER: First, with regard to Dr. Frank's point, we are talking about different replications. He is talking about replicability of findings. I am talking about replicability of the experiment. It is interesting to me that you can replicate findings without being able to replicate the experiment. This is a very thorny question. At what level of description does the behavior of the therapist have to be the same in order to consider the experiment replicated? This is a very tough one. It involves the analysis of variables such as commitment, warmth, permissiveness, and any others you may think of which may be affective even though other techniques or other aspects of behavior vary. So we are really talking about two somewhat different things.

The point I was trying to make about replicability was: how can your experiment be replicated when there is so little known about the behavior of your therapist?

DR. FRANK: The results could be replicated if the relevant variables have been adequately defined. In our study

the details of the therapist's behavior were assumed to be irrelevant variables.

DR. HAMBURG: What would constitute a replication of an experiment? Must every detail be the same?

DR. BUTLER: I have confessed ignorance but the general principle is that you are able to specify experimental conditions. I think there is a lot of meaning contained in that phrase.

DR. SASLOW: You always specify them to a certain degree; temperature, pressure, dimensions, and you can specify whether or not the therapist's behavior is going to be so and so. How do you get around Dr. Frank's point?

DR. BUTLER: Maybe you can't. Maybe I was simply saying in effect that perhaps there should be another level of specification. In regard to Dr. Wallerstein's discussion, I found myself in my comments agreeing in substance with him. From the simple-minded point of view of naturalistic observation and naturalistic inference, your study is much farther along the road toward experimentation than the sort of thing which I was talking about, which is much more simple. There are many more controls built into that than in the situation I discussed.

On the second point about astronomy and geology it seems to me I will hold firmly to my position that experimental method means from a scientific point of view the best means of knowing what you know. Astronomy and geology depend very much upon the experimentation done in the field of physics, chemistry, et cetera. Their controls have been provided for them by the positive knowledge of other sciences. Astronomy was, I believe, to a very large extent a macro scale elaboration of statistical mechanics until atomic theory was well developed. I think the same is true of geology. It was very much a descriptive science. I would not want to argue about the

meaning of science, but I think their controls have been given to them.

DR. SNYDER: I would like to address a question to Dr. Robbins and Dr. Wallerstein. One of your major independent variables is the therapy itself. As I see it, so far you don't measure this. How do you know what actually is taking place?

Perhaps this is just one of the things that Dr. Butler said or implied, or maybe I have not picked up the information; but we have so much evidence to indicate that therapists vary, and what they do varies so tremendously, that it means almost nothing just to call it a certain kind of therapy. Can you be sure that psychoanalysis is always so identical that it can be predicted absolutely because it is always of a certain form? In one study that we did, we had to throw out the findings of 40 per cent of the therapists, because they simply would not follow the rules and conduct the therapy they had been asked to conduct, even though they were specifically trained as to method. We were working with students and you are working with real therapists. Why can't you, with one of the best pools in the world of psychotherapeutic data, measure the actual therapy that goes on?

DR. ROBBINS: In the first place, I don't think the broad term of psychoanalysis does define therapy either. What we are more concerned with, and hence the list of variables under treatment—therapist variables—tries to describe all the components that we think are thus far relevant. There may be many more or others that could characterize or accurately describe the therapeutic process in which the therapist is engaged. We can then crudely characterize a therapist's position in relation to the different forms of therapy as being experimentally supportive, expressive, or what have you. Those final categories are unimportant. Somebody used the term habitat—I think

you did—I could not help wondering if you were referring to our concept of climate of therapy. What was the climate? What were the techniques that he used? What were the bases for determining variations in approaches and the like? So we are trying to describe a whole set of characteristics of what a person who calls himself a therapist does, both as technician and as a human being, in qualities of warmth and so on.

We feel, however, we have notions about therapy. We have notions about subtle differences between even my analytical approach versus Dr. Wallerstein's which have much in common. I think we probably have more in common than we have differences. Whether what works is determined by the things we happen to have in common or is determined by the things that are different, time will tell.

I would like to say a word while I think of it about this matter of the differences in styles of psychotherapy. I think it is extremely important that we keep this in mind. For instance, in Dr. Frank's paper, he talks about seasonal variations in feelings of patients. Obviously seasonal variations are not a concern of ours because they are with us several seasons several times. That is a lot of psychotherapy to a lot of you folks. It is not a lot to us. That is one of the big differences between all of us that I think we better lay on the table right now. Our use of the word psychotherapy is quite a different order of experiment to which some of these points you make are even more important.

DR. LUBORSKY: May I make an additional comment on the same answer? Dr. Robbins mentioned mainly the therapist's reconstruction of the type of treatment. I want to add that we have also the supervisor's reconstruction. A supervisor meets regularly with the therapist each week for an hour and discusses the

process of the treatment; we therefore can obtain well-informed views on the nature of the treatment. We also have the patient's reconstruction of what the therapist did. In more than half of our cases we also have the process notes that the therapist kept as he went along hour by hour. So that although we don't approach the fidelity of a recording, we have the possibility of conceptual validation from several points of view of what was done.

DR. GREENBLATT: My comment may not be consistent with the present trend of the discussion. It goes back to the problem of what seemed to me to be at one end of the scale the tight experimental design and at the other end of the scale, the freer naturalistic or correlated method.

I would like to point out that much of this is a function of training and personal preference. It so happens that experimental methods in psychiatry come largely from the allied field of experimental psychology where variables are fewer in number and perhaps easier to comprehend than those considered in clinical practice. However, I think we should not let the experimental psychological roots hold us back from gaining knowledge by newer or different methods.

We are on a voyage of discovery because the field of psychotherapy is extremely complex, as many people have pointed out. What is most appropriate to this voyage of discovery one asks? What paraphernalia do we need and what do we anticipate as to storms and distance our ship has to go. These are the important things and not any adherence to any specific method.

When a man brings me a set of data or conclusions, I find, as I suppose others do, that I react as a whole person to what he said. I would be inclined to reject a conclusion formulated within the best canons if it did not make any

sense to me. A priori there is nothing that I know of that tells us that one method of research is better than the other, especially in finding out something in a new terrain.

I think it is possible to speak of many varieties of methodology between experimental on the one hand and the correlative on the other. Many people during their lifetime are able to shuttle back from one to the other.

DR. ROGERS: I would like to pose a question to Dr. Frank and to the group which grows out of a very current interest of my own.

On this problem of controls I agree that there are difficulties with the own-control procedure and the equivalent-control procedure. I certainly feel there is much to be said for the notion you have used of having different therapists as your control in order to compare differential results. But as I have puzzled over this issue in connection with the research I hope to undertake in psychotherapy with schizophrenics, I phrase it this way. What type of therapy would be most sharply differentiated from the one I propose to use, and which therefore might be the best one to use as the control measure? If they were sharply different, we might learn more than if they were almost the same.

In that connection, I have wondered about the work in operant conditioning which springs from Skinner. It has been used in a way most relevant to psychotherapy by Ogden Lindsley in his work with schizophrenics, using a machine similar to a vending machine. The patient is placed in contact with this machine; it rewards him for different types of behavior, including a number of social forms of behavior. The thought has crossed my mind that if I could get Lindsley or some of his cohorts to deal with patients who are matched as closely as possible with the ones I would be

dealing with, it then seems to me there would be some very sharp differences which might make quite exciting findings.

On the one hand, I would hope that they would be willing to do their work in a manner which I think would fit their theory, with an absolute minimum of personal interaction, so that the patient is simply put in contact with this machine for an equivalent number of hours per week. The theory behind what they are doing is sharply different. I toss that out, at any rate, as one possible control notion. I am also raising the question whether in thinking of differential types of treatment we should try to maximize that difference so as to learn more sharply from the findings.

DR. FRANK: It seems to me to be a brilliant notion. I think that is what you need. The only control that you can be sure is really different is one which does not involve another person as therapeutic agent.

DR. ROGERS: That is really true. If you try anything else, say having a layman come in contact with this person for an equivalent number of hours a week, he may have very therapeutic attitudes. I have also thought—and this would run into more difficult professional problems—that I could pick some psychologists and/or psychiatrists whose attitudes are really anti-therapeutic, and put them in contact with a patient for an equivalent number of hours. But that runs various risks of real damage to individual egos. That is why this other notion struck me as being more feasible and perhaps equally significant.

DR. ROTTER: I would like to make a very brief comment on this. There is a third alternative in methodology which we have not discussed this morning, and that is the alternative of taking hypotheses out of therapy and bringing them into the laboratory. The notion that there is something either mysterious or

peculiar about psychotherapy is being discarded. If we are dealing with human behavior and try to take some of these notions into the laboratory, some of the difficulties of research might disappear.

DR. BORDIN. What has troubled me in research in psychotherapy is the question of to what extent are the data that I am collecting reflections of 1. how psychotherapists *think* and *talk* about psychotherapy, 2. the actual operations occurring in psychotherapy, and 3. my selection and interpretation of what has gone on?

This I think is apart from the issue of experimental control versus naturalistic observation. Clearly some of our concerns in psychology are analogous to the early worries of the astronomer regarding the extent to which he was generating his data and the extent to which the data were generating themselves apart from him.

This worries me a great deal. It results in my responding positively to Butler's notion of trying to develop our data collection in such a way that the observer operates minimally. The data should ideally be a function of what he has observed rather than what has gone on in his head as he has observed it.

I think those who presented the papers today and many of the rest of us try to evade this pressure with the argument that if you try to maximize what the data are doing and minimize the observer's operation on the data, you end up by destroying the data. How do you get significant observations while minimizing the internal operations of the observer which finally crank out the data? This is what I say is the problem.

The answer that Robbins and Wallerstein give us is: "Let us use the expert observer. Let us use him as an instrument and count on the fact that he is an expert observer to help us solve the problem." That does not satisfy me. My

problem is which group of expert observers to choose. What we know is that many highly respected clinicians, who are the expert observers, have gone through the observation and have come to different conclusions. I can't find this very satisfying.

I think I have outlined one of the dilemmas as I see them, and what I am interested in is to see if others are worried enough about this—I think they are—and the interaction we get on it to find a better solution.

DR. ROBBINS: Dr Bordin has hit a very important point that concerns me, and it concerns me not only in relation to our own work, but many of the other studies that are going to be presented here.

In the first place, you cannot observe everything. You have to make some decisions as to what you are going to observe, no matter how biased or unbiased you are going to be, or think you are. Immediately, when you decide, "We are going to observe these phenomena and not those phenomena," you have made a selection and the question to me is, what determines what you are going to observe. Usually I like to think that somewhere, explicit or implicit, conscious, preconscious or unconscious—I prefer the first two rather than unconscious—is a hypothesis or theory that the observer has used in selecting what he is going to observe. This already brings into question the circularity of the experiment. I think we kid ourselves in a sense when we say we are not going to be influenced by any bias. I think it is healthier, for better or worse, to say that these are the assumptions or these are to the best of my knowledge the bases for my having selected these particular phenomena to observe. I think this is a crucial issue to our whole discussion.

DR. WALLERSTEIN: I share Dr. Bordin's question, which is one of our central questions. He said the question is, how can you tell the difference between what goes on and what you think goes on, and what is reported as going on? He uses for comparison two situations: one in which you have a PGR and the reading on the face of a dial, which can be read with relative accuracy, and the other in which you have a response to a statement or question. What are you left with in evaluating that? He says it is a little unsatisfactory to be left with self-styled expert judges. This is the problem, I think, and the one we have to try to get around. The first step in trying to get around it is to ask what are these judges, expert or not, basing their judgment on? Can you force them to make systematic statements within pre-defined categories and in each case have enough relevant categories and get enough other judges to add their statements in these categories and get a consensus? Can you force explicit statements within each of these categories, and can you juxtapose them in ways that lead to predictions where you can see that the predictions stemmed from the assessment in these various categories? In that sense you have forced these experts a step further to define what their expertness is and to make it comparable with what other judges would say with regard to the same categories if they could agree on the value and importance of the categories.

I would like to go back to what Dr. Greenblatt said before about the discovery. I hope we are not taking an either/or position. Rather I would like to leave our position with his, one of looking at science not as defined by a specific method or body of methods, but rather as a set of problems to which you bring those methods which are most appropriate to it at any given stage of its development and not excluding a swing

from the most free wheeling kind of observation on the one hand, to the most rigorous experimentation if suitable to a particular problem.

DR. SASLOW: I would like to speak a little bit to Dr. Bordin's question, but approaching it peripherally to see if it throws any different light on it. It is possible to look on these two different approaches as being on some kind of continuum, of course. One could think from that point of view that Dr. Frank's approach is one conceived of in terms of successive approximations, starting with coarse, relatively easily definable variables. Even when he begins in this gross way, he turns up a number of questions which it is immediately clear have no obvious answer. One can think of at least two alternative explanations of such a simple thing, for example, as why minimal therapy patients did not do as well as those in the other two kinds of therapy. But it is possible to begin testing each of these hypotheses by further suitably designed observations which is the point which occurs to one as he listens to Dr. Frank's particular way of getting at the general problem. In this way, a solidly built body of knowledge could develop.

To take another question which his work has raised, why did it happen that more patients who came from lower social class levels were assigned to group therapy? If you look at the work of Redlich and Hollingshead, you have a strong suspicion that the therapists in the assigning phase may have decided that such patients don't fit the supposedly available criteria for individual therapy. Maybe they are not like this or not like that. I mention this as an example of the possibility of conceiving of alternate hypotheses which, if you start at the end of the continuum from which Dr. Frank did, may be thought of as a series of questions that require the designing of

experimental observations for meaningful answers to be found, and hence indicate the possibility of a solidly increasing body of knowledge. That is one way of looking at that part of the continuum, it seems to me.

Without going to the extreme which Dr. Rogers suggested in comparing diverse approaches to modifying human behavior, and while utilizing Dr. Frank's approach, one could, I think, take the same group of therapists and encourage one of them to emphasize a purely ahistorical approach, another one a purely ontogenetic approach, and another one a purely symptom-interested approach. As such refinements in their turn raised new questions which the experimenters could not resolve, there would be new experiments, further knowledge, etc. This I can easily see, as one moves from his end of the continuum towards a different point. If you look at the other presentations from this point of view, such as the Robbins-Wallerstein one, this is a totally different approach in which one attempts to answer many more questions simultaneously while the numerous questions themselves are asked at a number of levels which range from very simple to extremely complicated and subtle. By and large scientific experience indicates that the odds are against any easy or satisfactory answers being obtained in this way.

A second question pertinent to a point made by Dr. Wallerstein is this: How does one get outside the framework of one's own set of conceptions and obtain the necessary independent verification of what one is doing? You are dealing with a series of statements and observations made by persons presumably all of whom are being trained in the same framework of theory ("psychoanalytic") which Dr. Robbins has described himself as sharing. To what extent will it be possible for these particular therapists to conceive of alternative

ways of looking at the interactions they have with patients than the one with which they start? How does one get outside of this system is a question which leaves one rather uncomfortable. Even if there is the possibility of prediction in the idiosyncratic sense it is still necessary, for sound scientific work, to conceive of alternatives to one's initial hypotheses. I wonder how one gets around this difficulty.

DR. LACEY: Dr. Saslow has taken the words out of my mouth, but that won't prevent me from talking. In psychological circles today, we seem to be very satisfied, as scientists, if we can predict something, even if we lack detailed knowledge of what has gone into our "prediction." We sort of give up on the notion of deriving quantitative functions and laws and regularities, and a knowledge of variables. We say that if we can show that we can predict, then somehow we have acquired some status, that we have now some power. I think Dr. Saslow has put his finger directly on it. Our predictions are rarely of the sort that are decisive. As Dr. Butler said, they frequently impose upon us no restraints. In particular, I would like to ask Dr. Wallerstein and Dr. Robbins about the nature of their predictions, which are of this kind: if something happens, then something else will happen, *if*—and then the prediction is carefully qualified by a series of contingencies. I think Bavelas' subjects could have a heyday with that kind of contingent multiple prediction. I don't mean to be nasty. I am trying to point out that there seems to be great scientific danger, in our present state of knowledge, in having a sense of satisfaction that we think we can say we have predicted correctly.

DR. LUBORSKY: I would like to comment further on Dr. Bordin's comments. The question that bothered him

a great deal concerned observer validity, or how can one be sure that the observer is accurate. This was the problem that almost got the conscientious young astronomer fired, according to the famous story of the Greenwich observatory. This astronomer had to report that he did not see a natural event at the same moment as his superior did. We have been describing today the experiment at Menninger's as a naturalistic experiment. We followed the classical way that the astronomers developed for describing certain natural phenomena, that of getting measures of observer reliability. Perhaps we have not emphasized this as much as it should be in our report. Especially by the paired comparisons and independence of observer's judgment we have the possibility of estimating observer reliability. At the termination of treatment, for example, we have the patients ranked by the paired comparisons method, and each of the observers does this separately. These quantitative measures are additional to the detailed accounts in the various forms Dr. Wallerstein described, of how these judgments were arrived at. Certainly observer agreement is not the final answer, but I think it is the best way to get at whether the observer is seeing something out there.

So much for that point. I wanted to make some further comments and raise questions on Dr. Frank's paper. I very much agree that the minimal treatment group came out as it did because they got less treatment. This is in accord with good experience. I wanted to share some of the thoughts that I had which are mainly questions about alternative ways in which this result might be understood.

The inferior results from the minimal treatment could be attributed to differences in sampling of patients. The difficulties of matching patients makes the random assignment method very attractive. I like that direct way of settling the issue, by making a random assign-

ment. However, it is still possible that with 18 patients in each group you could very easily have very significant differences statistically on a sampling basis alone. I don't really know your data. You must have really compared them anyway at the end in terms of differences that came out.

DR. FRANK: Your point is still right. They matched on all the variables we thought of matching them on but there may be still important ones that we did not think of. I quite agree that with such small samples the possibility that the differences found may be due to sampling errors cannot be completely ruled out.

DR. LUBORSKY. On the third point, I think therapists cannot carry out different kinds of treatment with the same levels of skill. Furthermore, it probably requires unusual skill to be able to treat a patient adequately on a half-hour-every-two-weeks basis. I think treatment can sometimes be done on this schedule, but it requires more experience than other kinds of treatment require.

You mentioned as another possibility that minimal treatment was devalued both by the therapists and by the patients. I would like to say that this should be more than a possibility. Such a devaluation naturally occurs in a setting where longer treatments are given. This may not happen when it is typical in a setting to space appointments once every two weeks. It is not a question of whether such an atypical spacing had an effect,—it must have—but did this produce the result with regard to lack of effectiveness of minimal treatment?

Finally, a question about the superficiality of all three treatments. There is certainly a question whether with six months of treatment you can expect changes to occur in many patients. For some patients this is enough, but with many I would say that because of the

brevity of the treatment and inexperience of the therapist, the effects were not very great, perhaps on the order for some patients of what would be produced by a placebo. I doubt if this would be as true of long term psychotherapy.

DR. HUNT: I would like to ask a question about the procedures used in the research under way at the Menninger Clinic, I mean the procedures used to assure reliability of the clinician's judgments. Dr. Luborsky has just mentioned using the method of "paired comparisons." Does this mean that these comparisons are made by observers each of whom have interviewed the patients? Or, have the various observers looked at the same case reports? I gather the latter alternative must be true, and I ask this question because judgments of movement, made with the Movement Scale that we developed at the Community Service Society of New York, show a very high reliability when reliability is defined by inter-judge correlation. They show this high reliability if the two observers are looking at the same case records or summaries. For instance, in our field test of the Movement Scale, the correlation between judgments of movement from the caseworker on the case and those from an independent judge was $+ .82$. However, we hardly had a genuine measure of reliability here, for the independent judge looked at the case summaries that the caseworkers themselves wrote. Thus, the evidence that the independent judge got was that turned up by the caseworker.

While I am on my feet, I would also like to make some remarks about the method which Dr. Robbins and Dr. Wallerstein have described, and I would like to make these remarks apropos of Dr. Saslow's point that the hypotheses to be tested will be too complex to permit of test. As I read the paper, I found myself fascinated by what I conceive to

be the possibility of replication with certain of the predictive propositions in the Menninger program. I wish I knew more about these propositions. It is impossible to be sensibly critical about them without knowing them. On the other hand, it looks to me as if both the hypotheses and the predictive propositions might occur in classes which would occur repeatedly. I take it that these predictive propositions are of the "if, then" variety. If the same "if" conditions are tied to the same "then" consequences a number of times, and the observations fail to support the consequences predicted, it looks to me as if one could really discover that one was wrong and thereby learn something fundamental by this method. In case certain predictions from a given class are confirmed while others are denied, one would have a basis for looking for factors not taken into consideration, and again one might learn something by this means.

I say I am fascinated with this part of the Robbins-Wallerstein methodology, but I also have a "but." The "but" comes in in this way. Is the replication genuine or is it artifactual? Here is the basis for my question. It comes from the experimentally demonstrated effects of verbal reinforcement after the manner of Skinner's work with so-called operant reflexes. Last night, Joe Matarazzo was telling me that Dr. Silverburg has pointed out that his patients, when he was a Freudian psychoanalyst, produced Freudian memories and dreams. Later, when he was a member of the Horney school, they produced Horney-type memories and dreams. Still later, when he became a Sullivanian, his patients produced memories and dreams appropriate to the Sullivan picture of what is important in development. I believe this is an illustration of what William Verplanck has shown in laboratory-like situations will happen when a certain class of statements is reinforced by some response

of the person with whom the subject is in conversation.

What I am suggesting is this: If the clinician-predictor has his own propositions and his own expectations, is he by virtue of believing these things, making responses to his clients or patients which reinforce the kinds of patient-behavior which he expects? If he is reinforcing the kind of behavior he expects, albeit unwittingly, the substantiating behavior which he observes in his client or patient is but an artifactual substantiation of the proposition. This is a highly important issue as I see it

DR. WALLERSTEIN. The real puzzle to us from the beginning of our work is what kind of verification or support do we get outside the framework. This is Dr. Saslow's question and Dr. Lacey's, and I am sure the question of a great many others. I don't know. In a way I think of two somewhat contradictory positions in regard to that so I am sure both will be unsatisfactory. First, we can say we are not going to seek verification, it is not our problem. It is a problem of the field and it will be done in the sense that we operate within a given framework with its assumptions, from within which we set up our problems, set up our categories, generate our prediction, make our judgments. We hope that other people will set themselves to the same problems, will agree on similar categories as being important, and will make judgments and predictions in reference to psychotherapeutic events conceived within a different conceptual framework, that is, another psychological system. At some grand level, later on, the unity of science will be achieved when there are still others who bring the two together, like dropping one stencil on another, and seeing where they overlap and where they don't. That is one way of looking at it to me.

Second, we can accept the problem of verification as asking a lot but as being something legitimate. At other times I wonder if it isn't asking too much, in this sense. Does one ask of any other scientific endeavor that it be verified outside of its own framework? Can you validate microscopic phenomena other than by having somebody else look through the microscope? Is physics asked to validate its theories in extraphysical ways? I am not sure that the demand is made of others that is made within psychological science. This is unsatisfactory to me, and I am sure to others.

There is one other point with regard to Dr. Hunt's question concerning paired comparisons. We wish we could do it the way you say and that is to have different people have access to the *patients*. We are not able to. We are able to get different people with access to the same *data* and start from there to make independent judgments. As for your statement about the kind of replication we can get, certainly we have set up in each patient individually, predictions neatly tailored to his own status, and what we expect will go on in his individual treatment.

We also have the "core predictions." These are 70 statements on which, for each patient in turn, we make predictions and indicate which items will be true and which false upon outcome of treatment. In that sense we have an inter-patient correlation and can check to see whether the specific difference in outcome with regard to any one of these identical items of prediction is in terms of the variables we have set up. At least we hope we can with all the complexity of the variables and all the multiple kinds of teasing out that we have not yet really visualized how we are going to do.

DR. ROBBINS: I would like to enlarge on one point that Dr. Wallerstein covered somewhat in regard to the ques-

tions of Dr. Saslow and Dr. Lacey, and that is how to get outside of your own frame of reference. I agree with Dr. Wallerstein's comments because I feel we have two alternatives later. We are very conscious of the fact that we do have a built-in circularity, and this awareness I think is a help, although it is not a solution. This came to our awareness in a very simple way that I think is important to share with you.

Dr. Wallerstein and I sit down with the initial data and make a set of predictions. We do not share these. These are kept blind from Dr. Luborsky and his group at termination. We share a lot of other clinical and scientific issues together day in and day out so they can pretty well guess what we thought from the same data, even though they don't know precisely what we thought. We know if they stopped and said, I wonder what Wallerstein and Robbins said about this, they probably said thus and so, the probability is that they will be right. They can hear my uh-huh and uh-uh. We recognize this. There is no way of getting around it.

Suppose we find that certain hypotheses are worthwhile to be tested and we test them, and all within the circularity within our group, and there is the issue of alternative hypotheses. We have two alternatives. Should we test them or should somebody else? That they should be tested or that they should be re-examined from the point of view of alternate hypotheses I think is inevitable. That is the only basis for coming to a meeting where we know that people are operating from various alternate hypotheses. Our aim is to sit down and listen, and share the data as openly as possible with one another and see what happens.

This is why I think it is important to explicate as much as possible the inferential processes, starting with making your research assumptions as explicit as

you know how. My friends in physics say that physics after the point of the assumption is a very precise science. But the assumption is as vague as anything in psychiatry and psychoanalysis. It was from physicists and engineers that I got the courage to state assumptions with a degree of confidence, because they said the assumptions are seriously in doubt in physics. You don't know this because you are not physicists, but their assumptions are not any better grounded than ours. I think it is important to keep your assumptions explicit and share them. Respect the limitations of your assumptions as well as respecting the peculiarity of the assumptions of other people. I don't think you should try to get out of your frame of reference by yourself. I think you should share your frame of reference and listen to the criticism from another frame of reference. I think that is the only way you can do it in this very, very complicated field. That is the best I can do with this thing.

DR. FRANK: I would like to go along with what Dr. Saslow said. It occurred to me if the Menninger group is looking at psychotherapy through a low power microscope we are looking at it through the wrong end of the telescope. We are trying to discover the grossest and simplest aspects of treatment. It is interesting the way the discussion has gone. The questions directed to Drs. Wallerstein and Robbins have been connected with basic issues of philosophy of research. To me, they have been in the form of generating alternative hypotheses to explain the crude results which we have, all of which are testable, either by going back to the data or starting a new experiment.

For the record I ought to clarify two of those points, even though they are concerned with results rather than design. With respect to the possibility that more

lower class patients were assigned to group therapy because it was regarded as less desirable than individual therapy, if this factor were operating we would have expected that they would have been assigned to minimal therapy rather than group therapy. Actually lower class patients dropped out of minimal therapy less than group therapy. For them it was not devalued. To meet Dr. Luborsky's question about sampling, all you would have to do is do it over and see if the results hold up. There is one basic issue which will lead into other fields. Dr. Luborsky suggested that because of the limited period of treatment and the relative inexperience of the therapists our results might be superficial and might be of the magnitude of placebo effects. He seems to imply that the results of administering a placebo are small. We have been forced to the conclusion that placebo effects are not necessarily small. So the question of what is superficial and what is deep is really an open one.

DR. BUTLER: I simply want to reiterate my belief that in a field which is untitled empirically and scientifically, so to speak—or perhaps the fields—the distinction between experimental method and naturalistic approaches to observation inference disappears. It is obliterated by what to me is a fundamental fact, that a large number of alternative hypotheses are not excluded. So you come out generally with more hypotheses, and perhaps better ones. Then I would like to say, and this I think is something I did not make clear, but it is fundamental to my thesis, we have a fair philosophy of experimental inference which we can apply. I can apply it to Dr. Frank's study and come out with these rather crisp statements. But I cannot do that to Dr. Robbins' and Dr. Wallerstein's. I have to do what you do, which was to raise the alternative hypoth-

eses. I think what we need, quite aside from attaining this goal of becoming experimental, which I think is not too important in this context, is a similar theory of what you might call naturalistic inference which will allow us to apply the analogous criteria for examination of a study. There are rather formal considerations as well as these substantial questions about alternative hypotheses. I am trying to present in a rather small scale and elementary form an approach to a theory of clinical or naturalistic inferences. As far as choice is concerned, I think it is a matter of taste. If you want to come out with refined hypotheses and multiply them, I think naturalistic observations is a method par excellence.

I might say I felt I stood corrected by the conference on this matter of replicability because you can describe the conditions of the experiment at the level which Dr. Frank specified, and you can replicate on that basis. So it looked like some substantive considerations had crept into what I felt were formal considerations. I thought he should have described them more fully or with a different level of precision.

On the second point I am not so sure. What I mean by not so sure is that I am simply unsure. All the discussions I have read by philosophers of science specify or claim that specification of experimental conditions is necessary. It might well be that we have unearthed the problem for the philosophers of science. They have not said very well, at least to me, what it really means to specify a set of experimental conditions. It may be that those are really substantive rather than formal statements. They are not really philosophical statements. They have to do with alternative hypotheses instead. I am not so sure whether we have not unearthed the problem for them rather than for me.

The Dimensions and A Measure of the Process of Psychotherapy: A System for the Analysis of the Content of Clinical Evaluations and Patient-Therapist Verbalizations¹

TIMOTHY LEARY, PH.D., AND MERTON GILL, M.D.

Our research is directed towards developing means for the intensive detailed study of the process of psychotherapy.

When we set out we had a general idea of where we wanted to end up. We had two main goals. Our first was to develop a comprehensive set of dimensions for the description of the process of psychotherapy. Our plan for pursuing this goal arose from our conviction that the most meaningful statements one can get about a segment of psychotherapy are the evaluations of it made by well-trained clinicians. Clinicians—and when we use that term in this paper we mean a clinician evaluating a psychotherapy—however, make global, relatively unsystematic statements about psychotherapy, and certainly do not try explicitly in a clinical evaluation to cover all the dimensions of psychotherapy. Yet we did not wish to prescribe these dimensions to them beforehand, since this would have

obstructed our goal of developing a comprehensive set of dimensions. We decided therefore to have a group of clinicians—and there were twelve of us of mixed Sullivanian and Freudian training sitting in a circle around the recording machine—listen to tape recorded interviews and freely and independently write down what we felt was going on.

Working partly from the actual evaluations made by the clinicians and partly from a simplified form of the psychoanalytic model described by Rapaport (11) we evolved a model which covered everything our clinicians said in their evaluations. This model—which we later present in detail—combines two aspects which we may theoretically separate from one another. One is a view of the nature of psychological functioning—the constituents which go to make up this functioning and their interrelationships. The other is a view of the psychotherapeutic interaction—both intellectual and emotional—between patient and therapist and of how this interaction facilitates or hinders changes in the patient's psychological functioning in terms both of insight and clinical improvement.

We arranged the components of the model—constituents and relationships—into a group of what we labelled “clinical categories” and developed groups of variables for each category. The clinical categories thus became the comprehensive set of the dimensions of the process of psychotherapy. Now we could code any clinical evaluation, and count the entries in the various categories so we

1. This is a working paper, somewhat revised from the one presented at the Conference on Research in Psychotherapy sponsored by the American Psychological Association, April 10-12, 1958. The research on which the paper is based has been sponsored by the Kaiser Foundation of Oakland, California and supported in great part by Research Grants M-1323 and M-1323(C) from the National Institute of Mental Health, Public Health Service. The panel of clinicians who participated in discussions of the research and made the clinical evaluations included D. H. Powelson, M.D., Dorothy Bomberg, M.A., J. L. Dolhinow, M.D., C. Fernandez, M.D., J. E. Neighbor, M.D., Mary Rauch, Ph.D., S. Rauch, Ph.D., K. Schlesinger, M.D., A. Shapiro, Ph.D., and R. Suczek, Ph.D.

could compare clinical evaluations with each other. We called this a count of the content of clinical evaluations.

After having made a fair degree of headway with this first task, we turned to our second goal—to find some other way of analyzing the raw data which would give us some measurement of the psychotherapeutic process.

We wanted such a method because, important though we regarded clinicians' evaluations to be, we knew that even if they agreed, it might be simply because they shared certain general assumptions about psychotherapy and not because they were validly stating what was going on.

We wanted our second method to be quantitative and to serve two functions. First, it was to be a measurement of the process of psychotherapy, for the validity of which our clinical evaluation would be the criterion. Second, we would also use it in the reverse direction, to give us clues to the validity of the evaluation. We realize there are difficult conceptual and empirical problems in the use of the evaluation and the measurement to check on the validity of each other and we will later return to these problems.

An obvious possibility for a method of measurement suggested itself. Why not study the patient-therapist verbalizations in the same categories we had developed for the clinical evaluation? We knew that in insight oriented therapy—the kind which especially interested us—patient and therapist come progressively to talk more and more about the same things an evaluating clinician talks about. Would it not be useful to see if we could trace the progressive overlap of patient-therapist verbalizations and clinical evaluations both in terms of the kind of things discussed and whether or not they agreed in their specific assessment of these things?

Other people had studied the content of patient-therapist verbalizations in non-

clinical categories (9, 6) and there had been some studies which used some of what we called clinical categories (2), but there had been none which had used what we regarded as a comprehensive set of clinical categories, though Strupp's (13) system comes closest to doing so.² We felt we could regard ours as comprehensive because it had in significant part been derived from actual clinical evaluations and because it was developed on the basis of a theoretically comprehensive model.

Of course we found that our clinical categories covered sometimes more, sometimes less of the patient-therapist verbalizations. That part which was not covered we called "nonclinical." We decided we could more usefully study this nonclinical content if we made some kind of a systematic classification of it, so we devised a scheme for classifying all the nonclinical statements too.

To give the reader a more concrete idea of the detail with which we carried out this classification, we should say now that our unit for coding both clinical evaluations and patient-therapist content is the shortest verbalization which can be understood to be a combination of a subject—whether a person or impersonal—and some characteristic or attribute of that subject.

Of course we ran into many vexing problems in trying to get a reliable and adequately differentiated scheme for coding the clinical evaluations and the far more multitudinous things which a patient and therapist say to each other. In what follows we present the present state of our solutions of these problems.

2. We plan to discuss the relationship between Strupp's system and ours in detail in a later publication. We may note here that his categories are applied only to therapist verbalizations and that though his codings are made on microscopic units of content, they are ratings and hence clinical evaluations.

When we threatened to bog down in the minutiae of the details of our coding scheme we were urged on by the conviction that because our point of departure was something real and important about psychotherapy---what a clinician had to say about it---we were dealing with something significant and not just building an elaborate but meaningless scheme.

Our goal had now become the comparison of a count of the clinical evaluation of a segment of psychotherapy with a count of the verbal exchange between patient and therapist in that same segment to see: 1. How much of this exchange can be covered by the same categories which cover clinical evaluations? 2. To the extent that the same categories can be applied to both, how do they agree with each other? 3. Even when the same categories do not apply, will studies of the counts of clinical evaluation and of nonclinical patient-therapist content illuminate each other?

In the course of trying to answer these several questions we have developed a number of summary indices for both patient-therapist and clinical evaluation content, and in the hope that a foretaste of what is to come will help the reader to persevere in the unavoidably heavy going as we present our system we will describe here one of the most interesting indices which we will finally be in a position to calculate. It is an index of relevance---defining relevance as probable importance for the development of insight---which may be calculated for any verbatim segment of psychotherapy.

It is reached like this: by arranging our categories of subjects on one axis of a square and our categories of characteristics of these subjects on the other axis we have a matrix capable of classifying any and all statements in both clinical evaluation and patient-therapist content. Both subjects and characteristics can then be ordered on hierarchical continua in terms of what seems from the clinical

point of view the most important things to discuss in an insight oriented therapy. By so ordering them and by weighting the various segments of the matrix accordingly we devised our index of relevance.

Our presentation will follow the sequence: 1. The psychological-psychotherapeutic model derived from clinical evaluations and the psychoanalytic model and the clinical categories into which we organized the components of our model; 2. The nonclinical subjects and characteristics added to the clinical ones to form the matrix for classifying any and all statements in the two kinds of content; 3. A set of special clinical categories derived from clinical evaluations but not appearing as such in the model, 4. Several groups of modifiers of statements which were required to differentiate statements with obviously different clinical significance which would otherwise be coded as though they were the same; 5. Problems arising in the application of the system, and 6. Illustrations of the use of the system, including recording and tally sheets.

THE SYSTEM FOR CLASSIFYING PSYCHOTHERAPEUTIC INTERACTIONS

The Model

A study of the freely written comments by our clinicians on the psychotherapy interviews shows that they discuss in the main two kinds of topics. One kind has to do with the patient's psychological functioning and can be recognized as specific and concrete instances of what Rapaport (11) has described in general terms as the psychoanalytic model. The other has to do with the emotional and intellectual interactions between patient and therapist and how these either facilitate or hamper insight on the one hand, and therapeutic progress on the other. We have combined these two kinds of considerations to make a model which states in general

terms what is specifically discussed in clinical evaluations. The model is called a combined psychological and psychotherapeutic model. Its psychological aspect is a much simplified form of the psychoanalytic model as made explicit by Rapaport, in addition to which it includes intrapersonal communication between the topographic levels of the psyche—defining topographic in the psychoanalytic sense of relationship to consciousness (4). Its psychotherapeutic aspect deals with the interpersonal communication—both emotional and intellectual—between patient and therapist and how this facilitates or inhibits changes in the intrapersonal communication and in behavior (1).

We believe it important to point out that though the model is derived from the psychoanalytic one, it is implicit in the evaluations of most clinicians—not only those psychoanalytically trained.

The model takes its point of departure from motives, which therefore form one of its four principal constituents. The other three are drawn from the three possible fates of motives. In the conative model the fate of the motive is action, in the affective it is affect and in the cognitive it is thought. Clinicians talk not only about these four major constituents but they also describe three of them—motives, affects, and ideas—topographically, that is, whether they are conscious or unconscious, and they discuss whether

or not these three remain intrapsychic or reach overt behavioral expression—motives in action, affects in affect expression, and ideas in speech. We will call the unconscious, conscious, and overt behavioral expression the three levels of our model.

As we have already remarked, most of the statements the evaluating clinician makes are about clinical matters. We may now restate this by saying that the patient's ideas which the clinician will mainly discuss are similarly, the ideational representation in the patient of his motives, affects, and actions, or the patient's ideas about these three. Clinicians frequently attempt to specify how they have inferred the conscious and unconscious intrapsychic constituents—motives, affects, and ideas—from the overt behavior—action, affect expression, and speech.

The topographic position of the intrapsychic constituents of the model as well as whether or not they reach overt expression is a reflection of the play of forces in the mind, that is, of mental dynamics (12). Unconscious motives may be discharged into consciousness and then into action or they may be discharged directly from the unconscious into action. Ideas likewise may move from the unconscious into awareness and then into speech or directly from the unconscious into speech. As a constituent changes from one level to another it

TABLE 1

THE THREE LEVELS OF NONIDEATIONAL AND IDEATIONAL CONSTITUENTS

<i>Level</i>	<i>Nonideational Constituents (Motives, Affects, Actions)</i>	<i>Verbal Constituents (Ideas)</i>
Overt Expression	Non-verbal Behavior (Action, Affect-expression)	Overt verbal behavior, i.e., talk
Conscious	Conscious motives and affects	Conscious ideas
Unconscious	Unconscious motives and affects	Unconscious ideas

carries the same content, either directly or in distorted form. An unconscious angry impulse for example may become a conscious angry impulse or angry behavior. It may undergo various distortions, for example, reversal into a kindly conscious impulse or kindly behavior. Any constituent of the model which is being described as static, we will say is *at one of the levels*. Speaking loosely—in that the topographic position of the constituents is only a reflection of the mental dynamics—we may say that the constituents of the model can *move from one level to another*.

Movement between levels may take place either in the discharge direction—unconscious→conscious→overt expression—or in the reverse direction. What we may call a partial movement toward discharge may take place if the content of the constituent is altered—for example, if the hostility is expressed to someone other than the one for whom it was originally intended. Such partial movements as well as inhibitions of movements are conceptualized as the clinically familiar defense mechanisms. We distinguish two kinds of such defense mechanisms—the partial discharge made possible by an alteration of content, for example displacement or reversal, and general inhibition of movement towards discharge, for example repression, or the consequences of such general inhibition, for example thinking or talking superficially, confusedly, or evasively.

The therapist in insight oriented therapy—and the patient too insofar as he is in part cooperating—attempts by various maneuvers to diminish these inhibitions of discharge into consciousness and overt expression, whether action, affect expression, or speech. These maneuvers may be not only intellectual, like interpretation or clarification, but they may also be emotional behaviors, like friendliness or scolding. It must also be pointed out that discharge is not necessarily synony-

mous with insight—the acting out patient may be discharging impulses indiscriminately, but insight is being obstructed rather than facilitated—nor is it synonymous with therapeutic gain—the patient with extraordinary insight (or is it?) but no behavioral change is a familiar figure

Categories Derived from the Model

The model presents us with a hierarchy which we may designate as static condition (at a level), movement, cause of movement. It is from this hierarchy that we derive the first group of clinical categories to be used in coding. We number the categories in reverse order from the hierarchy just presented because we want our numbers to run from what we consider the most important category in bringing about or preventing change to the least. Our first category therefore is that of the intellectual interventions which bring about or hinder insight and admission into speech.³ Our second, third, and fourth categories are the three general kinds of movements which the model indicates—movement between conscious and unconscious, movement of nonideational constituents into overt expression, that is action and affect expression, and movement of ideas into overt expression, that is, speech. It must be remembered that these movements may take place either towards or away from discharge. Again we have on clinical grounds numbered these movements in terms of our view of their importance as a reflection of the facilitation or obstruction of insight—the lower the number the greater the importance. Our second category then is movement of ideational constituents into or away from speech, and is called “admission into speech”

3. We do not mean to imply that emotional interactions may not also play an important role in facilitating or obstructing insight and admission into speech.

Our third category is movement into or away from consciousness and is called "admission into awareness." Our fourth category is movement of nonideational constituents into overt expression and is called by the general term—"discharge," since by far the greater quantities of energy are discharged in this kind of movement.

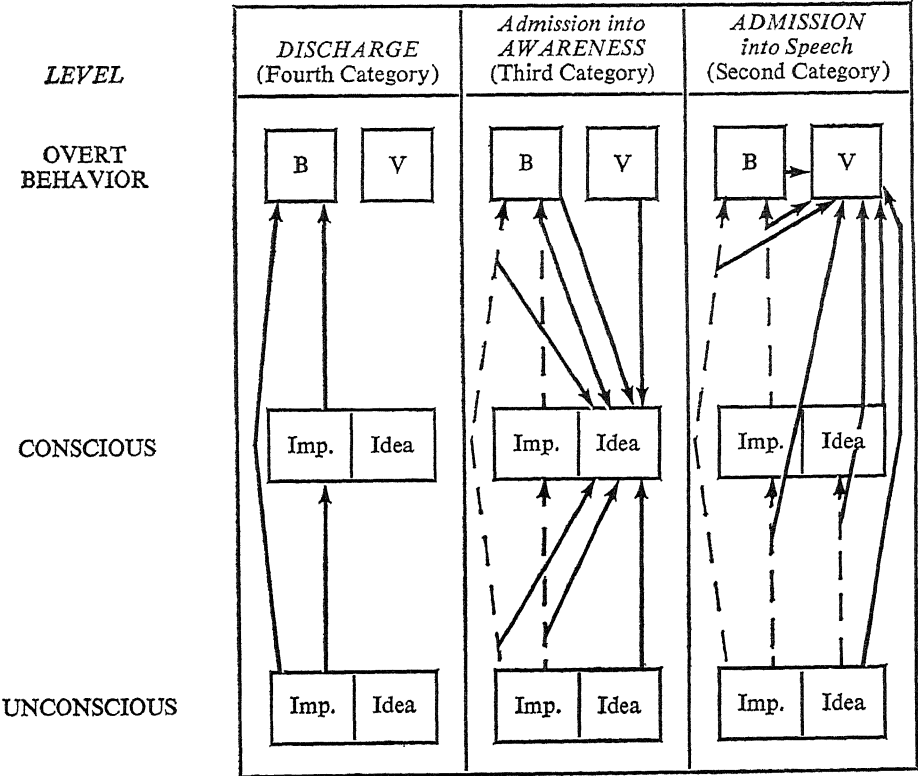
It has been emphasized that these movements reflect the dynamic play of forces in the mind, and that we are dealing with movements which occur against resistance. They are therefore coded only

when it is clear that the clinician is describing a movement taking place as a result of increased resistance (e.g., into unconsciousness of something previously conscious), despite resistance (e.g., reluctant admission into speech of something the patient has long known), or because resistance has been overcome (e.g., patient is finally able to express in behavior long inhibited feelings of tenderness).

The static condition, that is, the description of a constituent as being conscious or unconscious or of some overt

TABLE 2

MOVEMENTS TO OR AWAY FROM DISCHARGE, ADMISSION INTO AWARENESS AND ADMISSION INTO SPEECH (CATEGORIES FOUR, THREE, AND TWO OF THE SYSTEM)



Key: B = Behavior; V = Verbalization; Imp. = Impulses; Idea = Ideas
The dotted lines indicate processes from categories other than the one diagrammed. In the center diagram, for example, the dotted lines refer to Fourth Category variables (q.v.) which can be admitted to awareness or kept out of awareness.

behavior, without any indication of the forces at play in bringing about the presence of the constituent at the designated level, is our fifth category (e.g., the patient is behaving angrily; the patient unconsciously hates his brother). We will later make use of the fact that one can distinguish a particular content (e.g., the patient is hostile) and the level at which this content is described [e.g., the patient is behaving in a hostile manner, has conscious hostile feelings (affect), has conscious hostile wishes (motive), etc.] as a basis for designating the levels as the "dynamic" group of the "modifiers" to which we earlier referred.

We may summarize again that the fifth category designates a constituent at a level, the second, third and fourth categories movements between levels and the first category intellectual interventions which facilitate or hinder movements towards awareness and speech.

Under each of these categories we have a list of variables to permit a more differentiated analysis of content. For the first category, our variables are the different kinds of insight oriented interventions a therapist or patient can make, e.g., clarification and interpretation. For the second, third, and fourth categories our variables designate three specific

types of movement from one level to another: a) movement in which what is believed to be the original content of the constituent is unaltered, e.g., expression, and awareness; b) complete obstructions of movement, e.g., inhibition and denial; and c) "partial" movement made possible by altering some component of the original motive or affect, e.g., reversal and displacement. For the fifth category, that of the psychic constituent, we have three groups of variables—derived motive, action, or idea, bodily motive, action, or idea, and affect. Our distinction between "bodily" and "derived" is of course based on the theory—by no means now restricted to psychoanalysis—of a hierarchy of motives, with bodily impulses being the more primitive in the hierarchy (5). Our systems of variables have been worked out with varying degrees of care and completeness. The affect list is no more than illustrative and the bodily motives little more than that. The variables in the first four categories we regard as a fair beginning in the task of developing an adequate list. The derived motives list has been taken from the earlier work of one of us (8) and is more or less restricted to an "interpersonal" analysis. We shall not define these variables in this paper.

TABLE 3

THE FIVE CLINICAL CATEGORIES DERIVED FROM THE
PSYCHOLOGICAL-PSYCHOTHERAPEUTIC MODEL

<i>Number</i>	<i>Category</i>	<i>Example</i>
5	Constituent (ideational or non-ideational)	The patient is hostile.
4	Discharge	The patient inhibits her hostility
3	Admission into Awareness	The patient is unaware of her hostility.
2	Admission into Speech	The patient admits her feelings of hostility to the therapist.
1	Intellectual insight-oriented intervention	The therapist interprets her hostility.

TABLE 4
VARIABLES IN THE CLINICAL CATEGORIES

FIRST CATEGORY	SECOND CATEGORY	THIRD CATEGORY	FOURTH CATEGORY	FIFTH CATEGORY	
Insight Oriented Interventions	Processes of Admission, Denial or Alteration of Ideas in Verbalization	Processes of Awareness, Repression or Alteration of Ideas in Consciousness	Processes of Expression, Inhibition or Alteration of Impulses	Action, Speech, Motive, Affect and Idea	Affects
	Admit +, +-, - (The remaining variables are the same for 2nd and 3rd Categories)	Aware +, +-, -	Express +, +-, - Attenuate Alter Reverse Substitute Self-Substitute Other-Substitute	Bodily Motive, Action, Idea	Derived Motive, Action, Idea
				<i>Sexual</i> Sex (Un-specified) Orgasm Impotence Frigidity Stuck Retain Penetrate Incorporate Castrate Impregnate Show Look Manipulate Kiss <i>Aggressive</i> Beat (Un-specified) Bite Expel Rape Kill	A : Dominate B : Enhance self C : Compete D : Punish E : Attack F : Complain G : Distrust H : Derogate self I : Submit J : Admire K : Depend L : Cooperate M : Approve N : Support O : Nurture P : Teach ★ : Therapeutic neutrality
Structure therapy +, 0, - Question +, 0, - Focus +, 0, - Summarize +, 0, - Relate +, 0, - Interpret +, 0, -	Attenuate Alter Reverse Substitute Self-Substitute Other-Substitute Project Subject Project Object Repress Produce +, +-, - Think +, +-, - Internalize +, +-, - Emotionalize +, +-, - Deepen +, +-, - Clarify +, +-, - Process +, +-, - Structure +, +-, -				Calm Excited Depressed Elated Anxious Secure

* From the Kaiser Foundation Interpersonal System of personality diagnosis (8).
N.B. The codes, "+, 0, -", and "+-, -" modify the variables. The code "+, -" indicates the presence of the variable; "0" means the absence of the variable; "-, -" indicates the presence of the opposite of the variable or the inappropriate use of the variable; and "+-, -" indicates a less than direct presence of the variable (i.e., "indirectly, subtly, partially, almost," etc.).

The Non-clinical Categories

It will be remembered that we are seeking a way of classifying every statement consisting of a subject and characteristic and that both of these are to be hierarchically ordered in terms of their probable importance in producing insight. The great bulk of what clinicians evaluating a psychotherapy say is subsumed under the categories already presented. But much patient-therapist talk is not. As we stated in our introductory section we wanted to devise a classification of this "nonclinical" part of patient-therapist verbalizations. We settled not on one which we developed empirically but rather one which follows from man's imbeddedness in the biological, social and material worlds, one which the division of the sciences follows too. To take the subjects first: beginning with clinical and proceeding to nonclinical talk we order the subjects talked about by the patient, the therapist or the clinician, in increasing distance from what is important to the person whose statements are being coded, i.e., self in therapy, the other participant in the therapy, self outside of

therapy, people personally known to the speaker, people not personally known to the speaker, and impersonal talk dealing with matters psychological, biological (somatic), sociological, and physico-chemical (inanimate)

As for characteristics of subjects, after the five clinical categories already defined,⁴ the nonclinical characteristics follow the same sequence suggested for impersonal subjects: biological (somatic), social (for the personal subjects an individual in a social role), and physico-chemical or inanimate characteristics (rarely used for persons). Because a good deal of patient and therapist talk is often about social characteristics we include in Table 6 a further breakdown of this category

The relative ordering of the nonclinical characteristics is only a common sense one and it may very well be that their relative importance for the development

4. These first five are categories of characteristics since they are used to characterize persons as having the characteristics of these categories, for example, interpreting, repressing, displacing, hating, etc

TABLE 5
CATEGORIES FOR CLASSIFYING SUBJECTS OF A STATEMENT

<i>Code</i>	<i>Initial</i>	<i>Name</i>	<i>Example</i>
A	PT	The patient's behavior in therapy	The patient during the hour
	T	The therapist	
	PO	The patient outside of the therapy hour	Specify, e g., the patient's father
	PKP	People personally known to the speaker	
	NKP	Persons not personally known to the speaker	
	AN	Animals	Specify, e g., the patient's dog
B		Psychological subjects i.e., abstractions	Schizophrenia, aggression, Jungian therapy
C		Somatic Subjects	The central nervous system
D		Sociological Subjects	The role of a housewife
E		Inanimate Subjects	The patient's car

of insight actually has to be empirically determined for any particular case. Similarly we do not imply that there are regular or necessarily significant differences of importance between the various subdivisions of the social category.

Now that we have presented our clinical and nonclinical categories it will be readily recognized how the degree to which the clinical categories will cover what patient and therapist say can vary so widely.

It is common clinical experience that while at the beginning of an insight oriented therapy, patient and therapist talk little or not at all about the patient's *unconscious* motives, affects and ideas or about how the patient is defending himself against these or overtly expressing them, such talk may come to loom very large later in the therapy. But at the beginning, as well as later in periods of resistance, the patient may not talk even about his *conscious* motives, affects and

behavior. He may talk about cabbages and kings.

And of course in therapy which is not directed towards insight, because the therapist either in his general orientation or in the particular case does not regard insight as a desirable goal, the talk may remain what we have called "nonclinical" throughout the therapy.

The evaluating clinician on the other hand, no matter *what* patient and therapist talk about will discuss principally two kinds of things: a) the patient's personality and interaction with people important in his life; and b) the implications for insight and therapeutic result of the intellectual and emotional interactions between patient and therapist.

The Special Clinical Categories

But our matrix is not yet complete. Although any unit consisting of a subject and a characteristic of that subject can be subsumed under the categories already

TABLE 6

CATEGORIES FOR CLASSIFYING THE INANIMATE, SOCIAL, AND BIOLOGICAL CHARACTERISTICS OF PERSONS OR IMPERSONAL SUBJECTS

Number	Letter Code	Name	Example
11*		Inanimate (physico-chemical) characteristics	The lamp is lined with cork.
10	M	Miscellaneous behavior	The patient walked to the store.
	G	Geographic-temporal	The patient is a New Yorker.
	I	Intellectual-educational	The patient is a German major.
	P	Philosophic-religious	The patient went to church.
	S	Sociological-political	The patient is a Democrat.
	A	Aesthetic	The patient is a good painter.
	R	Recreational	The patient is a Giant fan.
9	O	Occupational-financial	The patient makes a good salary.
		Biological-somatic	The patient has a headache.

* Characteristics of this category are rarely attributed to persons.

* If at all possible, an action is scored in terms of its motive. If the patient "walked to the store" because she wanted the exercise the score would be 10 Recreational; if she had done so to express her anger at her husband the score would be in Fifth category "E" on the interpersonal circle. Ten Miscellaneous is used for an action where motive is unknown, or for an action in a social role which does not fit into one of the other social categories.

proposed, our study of our sample of clinical evaluations as well as our clinical experience shows that three categories are of sufficiently distinctive relevance to insight oriented psychotherapy to be pulled out of their general class and given a specific designation.

1. Psychological symptoms are so varied that any particular one could fall into almost any one of the categories already given. Anxiety, for example, could be fifth category, inhibition of affect expression, fourth, blushing, somatic, and kleptomania, social. Whatever a clinician designates as a *symptom* belongs in this special category, then, wherever else it might otherwise have been classified.

2. Statements about psychotherapy in the matrix so far presented would as subject be classified as a psychological topic or as characteristic be classified as social. If the characteristic falls in one of the first five clinical categories (e.g., nondirective therapy does not employ interpretation) it is classified in the appropriate category, but if it does not (e.g., I would like a change in appointment time) it is classified not under social (miscellaneous action) but under the special category for psychotherapy. This decision is of course an effort to provide a place for the special importance which usually attaches to talk about psychotherapy, whether in general or about the patient's own therapy.

3. Vocal (e.g., the patient's voice became very soft) and kinesic (e.g., the patient struck his fist on the table) behaviors would in the matrix so far pre-

sented come under the somatic category. These two kinds of behavior together with speech make up the three types of behavior which we as well as others (10) regard as constituting the interaction between patient and therapist. It is the special significance of vocal and kinesic cues as the basis from which the clinician infers at least some of his evaluations which led us to make a separate category for these two kinds of cues (only one category because they are infrequently mentioned).

We number these last three categories 6, 7, and 8, indicating that in relevance they stand between our first five clinical categories (with which we group them) and our nonclinical categories which we number 9, 10, and 11. Only with further empirical work will we know whether we need to add to or subtract from these three special clinical categories.

Our matrix is now complete. We repeat that it is employed to code statements expressed both in patient-therapist verbalizations and in clinical evaluations. While any particular statement from either the clinical evaluation or patient-therapist verbalization may appear anywhere in the matrix, clinical evaluation statements are likely to fall into the first eight categories, and patient-therapist verbalizations, much more frequently than clinical evaluations, will fall into categories 9 through 11. The logic of this matrix is such that (though only extensive experience can show how effectively and usefully it differentiates the material) it can subsume any verbaliza-

TABLE 7

THE SPECIAL CLINICAL CATEGORIES

Number	Name of Category	Example
8	Vocal-Kinesic	The patient speaks in a low voice.
7	Psychotherapeutic	The therapist is a Freudian.
6	Psychological symptomatic	The patient is phobic.

tion of patient, therapist, and evaluator. Our experience so far in using the matrix corroborates this logic. The complete matrix is diagrammed in Table 8.

The lower left doubly outlined area in Table 8 which in subjects extends up to "B" and in categories of characteristics extends up to "9" subsumes the clinical characteristics of persons and is the area

in which most clinical evaluation statements fall. The upper right hand box subsumes discussions about the nonclinical characteristics of impersonal subjects. The lower right hand box subsumes propositions in which the subject is impersonal while the characteristic is clinical. An example of such a proposition is "The sky is angry," which may be

TABLE 8

AREAS OF CONTENT DEFINED BY SUBJECT AND CHARACTERISTIC CONTINUA

			Categories of Subjects*											
			PT		T	PO		PKP	NKP		B	C	D	E
Categories of Characteristics	Physico-chemical	11	Non-clinical talk about people (the talk of ordinary social conversation)											
	Miscellaneous	10-M												
	Geographic-temporal	10-G												
	Intellectual-educat.	10-I												
	Philosophic-relig.	10-P												
	Sociological-political	10-S												
	Aesthetic	10-A												
	Recreational	10-R												
	Occupational-financial	10-O												
	Somatic	9												
	Vocal-Motor	8	Non-clinical		Non-clinical	Clinical		Abstract Psychological Talk Anthropo- morphic, metaphori- cal, autistic talk						
	Psychotherapy	7	clinical talk about patient in therapy and therapist		clinical talk (gossip or by patients early in therapy)		clinical talk about not known people							
	Psycholog. symptoms	6												
	Constituent	5												
	Expression	4												
	Admission into Awareness	3	Dynamic clinical talk about patient in therapy and therapist		Dynamic clinical talk about patient outside of therapy and personally known people									
	Admission into Speech	2												
	Intellectual insight-oriented	1												

* A key to the abbreviation of categories of subjects was presented in Table 5.

either anthropomorphic or psychotic depending on the context. It will be noted that this lower right hand box extends to the left only half way through "B" while the left half of "B" is labelled "abstract psychological talk." The reason is that the relation between psychological topics and clinical characteristics is different from the relation between somatic, social, and inanimate topics and clinical characteristics. The latter three kinds of statements are, as we have just said, either anthropomorphic or psychotic. And similarly, when a psychological topic has a clinical characteristic usually attributed to a person (e.g., psychosis rapes the intellect), the statement is likewise anthropomorphic or psychotic (the right half of B). But psychological topics can have as characteristics generalizations of the attributes of specific persons (e.g., certain neuroses are characterized by the inhibition of affect), and since such statements are neither anthropomorphic nor psychotic they are coded in the left half of B.

Modifiers

So far we have presented a matrix for classifying any statement, the greatest degree of differentiation in the matrix being in the area of statements in clinical categories. Up to now the description of our system has proceeded as though all statements were simple declarations. This is of course unjustified, because statements are qualified in many different ways which result in changes in clinical meaning sufficiently important to warrant the introduction of these modifiers into a scheme of classification. A statement may be presented for example as conditional, as a question, or dealing with something which occurred in a dream, as equivalent to some other statement, and with many other kinds of modifications. We will call these qualifications as a general class "modifiers."

Our first group of modifiers which we call "attitudinal" modifiers is a somewhat miscellaneous group which nevertheless bespeaks a particular attitude on the part of the subject toward the statement. It includes grammatical modes, e.g., conditional, interrogative and negative; evaluative attitudes, e.g., desirable and undesirable; and states of consciousness, e.g., dreams and fantasy.

Our second group of modifiers, those we call "dynamic" because they reflect the interaction of psychic forces, we referred to earlier and it has actually already been presented in our discussion of the "levels" at which a psychic constituent may appear and whether it is nonideational or ideational. An unconscious hostile impulse, the idea of a hostile impulse, an unconscious idea about a hostile impulse, a conscious feeling of hostility, hostile behavior, etc., are all about the same psychic content, hostility, but differ in whether they deal with motive or idea and whether the constituent is unconscious, in awareness, or overtly expressed. These various ways in which a constituent may be modified are the dynamic modifiers.

Modifiers in our third group are various kinds of connections between propositions, (e.g., causal, comparative, equivalent). These modifiers differ from the first and second groups in that they are not a modification of a statement in itself but rather are a specification of a statement in relation to another statement. We divide connections into three kinds, simple, causal, and change-improvement. The particular modifiers we suggest and even their grouping is still very tentative.⁵

5. We will later note that a modifier may be used by a speaker either to modify his entire statement, e.g., "I dreamed that . . ." or it may be a modification of something within the statement, e.g., "The patient is unconsciously hostile."

TABLE 9

THE MODIFIERS BY WHICH STATEMENTS ARE QUALIFIED

<i>Code</i>	<i>Description</i>	<i>Example</i>
<i>A. Attitudinal Modifiers</i>		
?	Interrogation	Is he angry?
z	Conditional	If he were angry .
!	Imperative	Be angry!
B	Negative value	His anger is a bad thing.
G	Positive value	It is good that he is angry
W	Wish	He wishes to be angry.
F	Fear	I fear he is angry
N	Negative	He is not angry.
Dr	Dream qualification	I dreamed he is angry.
Fan	Fantasy qualification	In my fantasy he is angry.
Iv	Involuntary	He couldn't help being angry.
Sig	Significant or important	His anger is important.
<i>B. Dynamic Modifiers</i>		
No code	Nonverbal overt action	He behaves angrily.
V	Talk or verbalization	He said he is angry.
C	Conscious ideation	He thinks he is angry.
△	Conscious affect	He feels angry.
○	Unconscious impulse	He is unconsciously angry.
UN	Unconscious ideation	He has the unconscious image of himself as angry.
<i>C. Connective Modifiers</i>		
(Simple connections, subdivided into four variables)		
con	Connection	His anger is connected with . . .
equ	Equivalence	His anger is the same as . . .
com	Comparison	He is more angry than hurt.
	(of two characteristics)	
cnf	Conflict	His anger makes it impossible for him to feel his . . .
(Causal connections, subdivided into two variables)		
S	Stimulate	His anger is an attempt to get . . .
C & E	Cause and Effect	His anger causes him . . .
(Change-improvement connections,* subdivided into six variables)		
wor	Worse	His anger is worse.
N ch	No change	His anger has not changed.
imp	Improved	His anger is less of a problem.
gro	Growth	He expresses his anger more maturely.
reg	Regression	He expresses his anger more childishly.
<	More than (temporal)	His anger is increased.
>	Less than (temporal)	His anger is less.

* Implicit in all change-improvement connections is variance over time (temporal change); the first five listed are more specific and take scoring precedence over the purely temporal connections.

PROBLEMS OF APPLICATION

Units of Evaluation, Summary Analysis, and Count

The division of the process of psychotherapy into units poses a problem which every researcher in this field has had to face. It must be noted that there are three different kinds of units needed: 1. the segment of material to be given the clinician on which he is to make his evaluation, 2. the segment of material to be used for summary analysis of count of patient-therapist verbalizations, and 3. the segment to be used in making the content count.

The natural unit of psychotherapy for the clinician is the session. But psychotherapy can be divided into units which for various research purposes may be more useful than the session. Fifty-minute units are too long for a clear delineation of the many events which take place in much shorter time spans. And fifty minutes is too short to encompass some of the longer and more slowly developing trends.

Basically the choice in process research is between arbitrary units or meaningful units. Arbitrary units are temporal (minutes, sessions, blocks of sessions), grammatical (the word, the clause, the sentence), or typographical (the page).⁸ Meaningful units require a proper definition of a critical event, and ways of determining the beginning and ending of such an event.

Meaningful units seem a rational division of the process, but arbitrary units can be more easily and reliably determined. In meaningful units it is often impossible to obtain agreement as to when a unit begins or ends. And the fact that there are many levels and cate-

gories in which events occur makes it difficult to determine which is the critical variable. Units determined in terms of interpersonal interaction for example may be very different from units determined in terms of intellectual activity.

We have not yet found a satisfactory way of reaching good interjudge reliability in classifying meaningful units. Reliable classification has been demonstrated by others (3) for units defined according to manifest topic or theme but this is successful only when an interview is broken up into only a few large units.

It may be desirable for some research purposes to have the same material divided into units of various length so that one gets clinical evaluations of short-term units and of the same units seen in the context of a long-term trend. A possible series of useful units would be: a) ten-minute intervals, b) single sessions, c) blocks of sessions, d) the entire therapy.

In our own research thus far we have asked the clinicians to evaluate consecutive ten-minute segments of an interview and then to make a summary evaluation of the interview as a whole.

The considerations advanced for *units for clinical evaluation* apply also to the *unit for the summary analysis of counts of patient-therapist content*. After the statements have been coded and counted, into what temporal units will they be combined for summary and tally analysis? We have settled on the ten-minute unit so that our patient-therapist content count summaries cover the same unit which the clinician is evaluating.

Turning to the issue of the unit for *counting*, we have already indicated our conclusion. Since coding of clinical evaluations or of patient-therapist content should be such that it can be executed by technicians and because the coding must be as objective as possible the unit should be such as to involve the least amount of inference. Our unit is the

6. The question of whether the speech (defined as beginning when one participant speaks and ending when the other does) is an arbitrary or meaningful unit is complex and we shall not deal with it here.

statement—already defined as the shortest verbalization which can be understood to consist of a subject and a characteristic.⁷

Sequential versus Randomized Units

It is possible to give clinicians or technicians the interview material in its original sequence or in randomized order. The latter approach is necessary only when the research design aims at preventing the listener from knowing what section of the therapy is being considered.

Where it is the intention to have the clinicians evaluate interview material as they do in their usual clinical work, it is essential that they be exposed to the raw data in the original sequence.

Since technicians simply code the content they can be given data in nonsequential order where this is convenient.

Reliability of the Coding Procedures

How easily and readily can this matrix be used by coders? The answer is given in the comparison of codings by coders working independently—that is, by reliability studies. Both the coding of clinical evaluations and of patient-therapist content must be tested for their reliability.

We calculate reliability separately for the several judgments involved in any coding—judgments as to scoreable items, subject category, characteristic category, level (for fifth category characteristics only), variable and modifier.

Table 10 presents the percentages of agreement in coding clinical evaluations between the codings of a judge (one of us) which were used as the criterion and

the codings by technicians working independently. The rules for determining “agreement-nonagreement” were rigidly defined. Failure to score the event or adding an extra score were counted as misses. For decisions as to category level, the non-fifth category variables the code was either “right or wrong”—with no half scores or near misses credited. On fifth category interpersonal variables, for which as we noted in Table 4 the Kaiser interpersonal circle was used, scores in the same octant of the circle were considered hits—all others misses. In determining the reliability of the coding of clinician evaluation, agreement as to subject was not calculated because the clinician designation of subject is almost always obvious; and modifiers, being relatively few, were lumped with characteristic categories.

The percentages of agreement range from 77 to 99. The coder’s reliability changes from clinician to clinician and from unit to unit. Some clinician’s evaluations are more difficult to rate because of the vagueness or complexity of their language. In the future when we publish analyses of interviews we plan to present agreement percentages for each unit and for each clinician so that the reliability of each set of data will be known.

Turning to the reliability of the coding of patient-therapist content, we report only one set of figures, based on the coding of 584 statements in one fifty-minute session. Agreement as to subject category was 99%; characteristic category 87%; and for variable, 97%.

The agreement percentages presented in this paper were calculated during our preliminary trials of the system. Our final analyses of the reliability of coding will be based not on the single scores but on indices—some of which will be described below—obtained from tally sheets on which individual scores are summarized for segments of therapy.

7. We have not yet worked out exact rules for determining what constitutes a statement, but we have found that even in the absence of such rules agreement among coders is high. The general rule is to code as a statement any and every verbal grouping which can be understood as a subject with its characteristic.

TABLE 10
PERCENTAGE OF AGREEMENT WITH FINAL JUDGMENT OBTAINED BY FOUR CODERS ON
FOUR CLASSES OF DECISIONS IN CODING CLINICAL EVALUATIONS

Coder	Types of Coding Decisions												Number of Coding Decisions		
	Codeable Event			Category			Level			Variable					
	1st Study	2nd Study	3rd Study	1st Study	2nd Study	3rd Study	1st Study	2nd Study	3rd Study	1st Study	2nd Study	3rd Study	1st Study	2nd Study	3rd Study
D	87	92	—	93	89	—	99	87	—	85	91	—	572	544	2777
C	84	83	87	84	90	88	87	84	96	88	77	95			
H	86	93	93	83	92	92	96	83	90	87	86	92			
J	80	90	86	86	89	88	91	77	96	77	83	85			
Average of all Coders	84	90	89	88	90	91	93	84	91	82	85	90			

The Problem of Obtaining the Clinical Evaluation

We have said that our clinicians evaluated freely, that is, they were asked to say whatever they wanted to in their own words—and that is what clinicians prefer to do.

If the clinician is permitted free evaluation of the interview material three major difficulties arise: 1. He may make complex, sometimes even obscure statements which are difficult to translate into the categories of the matrix because one may not be sure what he means, 2. He may say nothing about various areas, so that for this particular clinician no evaluation is available for this particular area, and 3. Different clinicians habitually evaluate at different degrees of depth—which also means with correspondingly differing degrees of inference.

On the other hand, if the clinician is restricted he complains that he cannot really convey his understanding of the process and must make forced choices. A compromise clearly must be struck between freedom and restriction and this compromise must be of such a nature that it clearly specifies the degree of inference to be used and requires a *comprehensive* description of the psychotherapy.

It was useful to us to get free descriptions since our main emphasis has been on developing the system of categories, but we found ourselves handicapped in comparing the clinical evaluations with each other because the evaluations were not inclusive enough. (It will be noted, however, that allowing free evaluation yields data for comparing clinicians with each other and the same clinician with himself at different times in terms of what the clinician chooses to discuss.)

The problem of the reliability of the clinical evaluation is clearly the issue here and it is that to which we turn.

Reliability of Clinical Evaluation

As any clinician will be quick to say, there must be a study of whether clinicians agree with each other in their evaluations. But a clinician might not be nearly so ready to admit that if clinicians were to agree in their evaluations this would be no more than a measure of reliability.⁵ He would be likely to feel that since clinicians seem so frequently to disagree, a demonstration of agreement would practically amount to a measure of validity—to the conclusion that the clinicians are stating what is really going on! But is this really true? Those who have little esteem for the insight of a clinician would of course say that it is not true but merely demonstrates that a particular group of clinicians have been so selected that they agree with each other. Those who esteem the insight of the clinician and know how frequently clinicians disagree would be heartened by a demonstration of clinician agreement but would nevertheless have to admit that, strictly speaking, nothing more than reliability could be claimed by such a demonstration.

A satisfactory demonstration of reliability of the clinical evaluations is of course essential before one can place much confidence in them in using them for comparison with a measurement of the raw data of psychotherapy.

With the free evaluations which we used, even the study of intra-clinician reliability by having the clinician repeat

8. Whether we may speak of "reliability" or only "agreement" in talking about clinical evaluations depends on whether we restrict the term "reliability" to the study of a measuring instrument and in turn whether we restrict the term "measuring" to a quantitative instrument. Our position is that the evaluation deserves to be called a measurement (though in this paper we mean our quantitative study of the patient-therapist verbalizations when we say measurement), even though it is qualitative (7, p. 118).

his evaluation after a lapse of time would be difficult since he might elect in the two evaluations to talk about different areas of the interaction.

We have nevertheless made some studies of interclinician reliability by comparing the pooled evaluations of the group split in half. Because our freely obtained clinical evaluations are not comprehensive, we use pooled clinical evaluations as our basis for comparison with counts of patient-therapist content. But it is clear that there are many ways in which such a method could lead one astray. We are banking on the assumption that the clinicians generally agree, that some of their apparent differences actually complement one another because they are made at differing "depth," and that idiosyncratic evaluations will not carry much weight in comparison with the mass of pooled evaluations.

Validity of the Method of Analysis

How can we be sure that our system studies what it is supposed to study—that it is valid? How do we know it measures what "really" goes on in psychotherapy?

It is exactly in order to be able to make a step forward towards answering this question that we have laid so much stress on the desirability of studying the raw data of psychotherapy by two different methods—the clinical evaluation and the count of the patient-therapist verbalizations.

We must nevertheless admit that the clarification of what constitutes validity in this research, let alone what methods may be used to demonstrate validity are problems for the future. Actually the methods for demonstrating validity are the same as the methods to be devised for comparing our two kinds of study of the raw data, since each kind of analysis of the raw data serves as a test of the validity of the other.

Yet from the point of view of establishing validity, our two methods differ from one another. The clinicians' evaluation is a global qualitative statement—though when we count its content we do state it in numbers—while the count of the patient-therapist content is much more akin to the usual instrument for psychological measurement. Our task would of course be much simpler if we could say that the clinical evaluation is the criterion of the validity of the patient-therapist verbalization measure. But since we cannot assume that the clinical evaluation is valid, we must also take the patient-therapist verbalization measure as a criterion of the validity of the evaluation.

In our later illustration of the comparison of the two kinds of analyses of the raw data of psychotherapy we will exemplify the several kinds of situations which arise.

It seems to us likely that as various methods of comparing the two kinds of analyses with each other are tried out the matter of the length of the units to be compared will begin to loom large. It may be, for example, that clinical evaluations will come to agree more as a therapy progresses and that—on the assumption that this means the clinicians are more validly stating what is going on—these later evaluations will show more consistent results when compared with the count of patient-therapist content. In other words it may be that to demonstrate validity one will have to work with long term trends.

A general rule which we suspect it will be useful to abide by in seeking ways to compare the two methods of analysis is that if the clinicians consistently report something, a hunt should be made for something in the patient-therapist count which runs parallel with it.

We recognize how inconclusive and incomplete these last remarks have been and stress again that the development of

useful methods of comparing the two kinds of analysis of the raw data is a task for the future. It may be that a really satisfactory measure of the raw data of psychotherapy cannot limit itself to the patient-therapist verbalizations but will also have to count vocal (voice quality) and kinesic (gestural) behaviors, which, as we have already remarked, together with speech make up the three kinds of actions into which psychotherapeutic interaction can be divided.

ILLUSTRATIONS OF THE USE OF THE SYSTEM

The burden of this paper has been to point up the desirability and feasibility of comparison between the count of clinical evaluation and of patient-therapist content in clinical and non-clinical categories, and the several parts of this following section are to lay the groundwork for an example of such a comparison. We will first illustrate unitizing and coding with examples of the original protocols and the recording sheets for the coding of these, then present tally sheets and summary analysis, first of clinical evaluation, and then of patient-therapist content, and finally compare the two.

We would first like to note however that besides the final comparison of evaluation and patient-therapist content a number of possibly useful comparisons within the evaluation and within the patient-therapist content suggest themselves. We may list them as follows:

1. Inpatient: patient's content for one unit compared with his content for a later unit.
2. Interpatient: one patient's content vs. another patient's content.
3. Intratherapist: therapist's content at two points in time.
4. Patient vs. therapist: patient's content for one unit compared with therapist's for same unit.

5. Intraclinician: a clinician's comments at two points in time.

6. Interclinician: one clinician's comments vs. another's (clinicians of same or different theoretical orientations).

In the course of this section we will incidentally present some examples of patient vs. therapist, inpatient, intra-therapist, intraclinician comparisons (this last for pooled clinicians).

Unitizing and Coding a Clinical Evaluation

We first present some of the practical working devices we have evolved—our recording and tally sheets—with the scores assigned to each statement in a sample of a clinical evaluation.

Table 11 gives a sample clinical evaluation of a ten-minute unit of psychotherapy. The unit is the third from the first session with this patient. Table 12 illustrates the scoring procedure, presenting the scores for each numbered unit, designating in order the speaker, unit number, speaker modifier, the subject of the statement, the category into which the subject falls, the object, the score (variable and what it is about), the modifier of the score⁹ and the category of the characteristic.

For Unit #1 "Therapist moves abruptly to requesting information on Patient's difficulty," the recording sheet reads: speaker, clinician No. 1; statement number, 1; speaker modifier, none; subject, therapist; subject category, A; object, none; variable, question (the intellectual intervention is that of questioning) and the object of the variable is symptoms since this is what the therapist has inquired about; modifier of the score, none; category of the characteristic, 1 (an intellectual intervention). And so on.

9. It will be easier to show the difference between the two kinds of modifiers when we demonstrate the scoring of the therapy protocol.

Unitizing and Scoring a Therapy Protocol

Table 13 contains an excerpt from a therapy protocol. It is from the same unit as the evaluation just scored but includes only the first minute or two of the unit. Each subject-characteristic unit is numbered. The same scoring procedure and recording sheets are used for scoring patient-therapist content as are used for scoring and recording clinicians' evaluations.

We observe that the first statement made by the therapist (#275) concerns psychological symptoms of the patient. The recording sheet shows in order: speaker, the therapist; unit number, 275; speaker modifier, W, which means the entire statement to follow is something the speaker wishes; subject, the patient; subject category, A (see code in Table 5); object, there is none; score, there is none, since there is no variable; score modifier, V_{pt} , which means the patient

is to verbalize¹⁰ (a subscript is used to indicate the person to whom a modifier applies unless it is the speaker's modifier); and the category of characteristic, 6. The entire code then means that the therapist has said he wants the patient to talk about her symptoms. The patient's reply (#276) repeats the therapist's statement in the form of a question. The recording sheet therefore reads: speaker, patient; unit number, 276; speaker modifier, question; subject, patient; subject category, A; score, no variable; score modifier, V as in 275 but no subscript since the speaker is the patient and W_{Th} , indicating that the therapist's

10. The difference between the two kinds of modifiers was discussed earlier. "Speaker modifier" actually includes both kinds: a) a modification by the speaker of the entire statement to follow, e.g., #280, "I don't know if Miss Y told you . . ." and b) a modifier referring to the speaker as part of the statement, e.g. #291, "for a long time we thought . . ." It is practical to group the two kinds of speaker modifier together.

TABLE 11

CLINICIAN'S EVALUATION OF THIRD UNIT OF FIRST PSYCHOTHERAPY SESSION

1	2
T. moves abruptly to requesting information on P's difficulty./	P. momentarily shocked./
3	4
Embarrassed/—regains poise,/	5
6	7
becomes intellectual/ and controlled talking about sexual	8
life./	9
10	11
Tries to be calm/ but occasionally embarrassment comes through./	P's defensiveness
12	13
more obvious as rationalization./	14
15	16
T. tends to encourage/	P. to return to generalization./
17	18
P's manner subtly/ suggests being willing to cooperate/	19
20	21
but will not express/ need./	22
23	24
Doubts that any man can help her sexually/ or therapeutically./	25
26	27
Covertly P. seems to move/ between embarrassed response/	28
29	30
to influence of T./ and	31
32	33
quick return to self contained/—and impervious manner/—also controlling./	34

wish is now part of the score; and characteristic category, 6. Then the therapist says (#277) "I know a little bit about it." The coder takes this to mean "I have an idea" and the recording sheet therefore reads: speaker, therapist; unit number, 277; speaker modifier, none; subject, therapist; subject category, A; score, no variable; score modifier, C, which is the code for conscious idea; characteristic category, 6, because the idea refers to

the patient's symptom (note that the category in an instance in which the entire score is a modifier is what the modifier is about, in this case, the patient's symptoms). And so on.

Tallying and Analyzing the Evaluations

We will first describe the instrument we use, then review the scores in our example and finally suggest the inferences which may be drawn.

TABLE 12

SCORING OF CLINICIAN'S EVALUATION OF THIRD UNIT OF FIRST PSYCHOTHERAPY SESSION

			Classification of Subject			Classification of Characteristic		
Speaker	State- ment Num- ber	Speaker Modi- fier	Sub- ject	Cate- gory	Ob- ject	Variable and its Object	Score Modifier	Category
Clinician								
No. 1	1		TH	A		(Symptom) Question		1
"	2		PT	A	Th	Anx		5
"	3		PT	A	Th	H ⁺		5
"	4		PT	A	Th	Sec		5
"	5		PT	A		Sex V Emo.-(Int)		2
"	6		PT	A		Sex V Emo.-(Int)		2
"	7		PT	A	Th	Calm	W _{pt}	5
"	8		PT	A		PT ^{H+} B Exp.-		4
"	9		PT	A		Uns V Emo.-(Rat)		2
"	10		TH	A		Emo.-	Stim.	1
"	11		PT	A		Uns V Spec.-		2
"	12		PT	A		I [*] B Exp. ⁺⁻		4
"	13		PT	A	Th	I [*]		5
"	14		PT	A		K [*] B Exp.-		4
"	15		PT	A	Th	K ⁺		5
"	16		PO	A	Men	G* (in sex)		5
"	17		PO	A	Men	G* (in therapy)		5
"	18		PT	A		H [*] B Exp ⁺⁻	Effect	4
"	19		PT	A	Th	H [*]		5
"	20		TH	A			Cause	-
"	21		PT	A	Th	B*		5
"	22		PT	A	Th	B*		5
"	23		PT	A	Th	A*		5

* Lettered scores from the Interpersonal System of Personality. (Sec Table 4).

TABLE 13

EXCERPT OF PATIENT-THERAPIST CONTENT FROM THIRD UNIT
OF FIRST THERAPY SESSION

-
- 275
- T. 121: Mmmhnn./ And tell me what brings you here? Hmm?/
- 276
- P. 121: (half-laugh) Go . . . tell this again?/ (laughs)
- 277 278
- T. 122: I know a little bit about it,/ but I'd . . . I'd rather hear it as you would put it to me yourself./
- 279 280
- P. 122: We first . . . ah . . . ah . . . had gone to see Dr. X./ I don't know if Miss Y. told
281
you or not/ . . . Jim had had two courses in . . . ah . . . one in educational psychol-
282
ogy and the other a psychology seminar./ And I had sat in on a number of the
283
sessions/ and in addition had had the two Dr. X's children in school./ and . . . and
284 285
liked them so well and/ became just acquainted with the parents./ And, so, when
286 287
this came up we first thought to consult him/ because he was the only one we
288 289
know./ And . . . ah . . . he in turn called Miss Y./ knowing that his schedule
290
couldn't possibly . . . ah . . . take us,/ and also that we couldn't possibly afford to
291
go to his regular practice./ Ah . . . for a long time we thought perhaps it was
292
just a . . . a physical barrier/ and tried to accept it that way/ . . . the fact that I
293
wasn't getting any satisfaction in our sexual relations,/ and then . . . ah . . . Oh,
294
within the last year, I expect, I just don't quite remember roughly . . . when it
295
began . . . ah/ it seemed like . . . ah . . . it was so unsatisfactory/ that it would
296
result in a . . . Oh . . . just . . . ah . . . woul . . . it would seem, for instance the
297 298
next day, that we would both be a little irritated more than usual/ because we
aren't that way often./ And . . . ah . . . mmm . . . it just gradually got to the place
299 300
where rather than . . . ah . . . do it/ or make any attempt/ we would often just . . .
301
just not do it all,/ thinking that it (half laugh) . . . the results after that weren't
pleasant either./
-

DESCRIPTION OF THE TALLY SHEET

The written descriptions of four clinicians who listened to tape recordings of the first and third units of the first session were classified by technicians and the results tallied in Tables 15 and 16. (These units were chosen for our illustration because of the striking difference between them.) The scores for the four

clinicians are pooled.¹¹ These clinical tally sheets are divided left and right into the emotional and intellectual interactions. Overt behavior, awareness, and the unconscious are represented on the left by the top, middle, and bottom thirds.

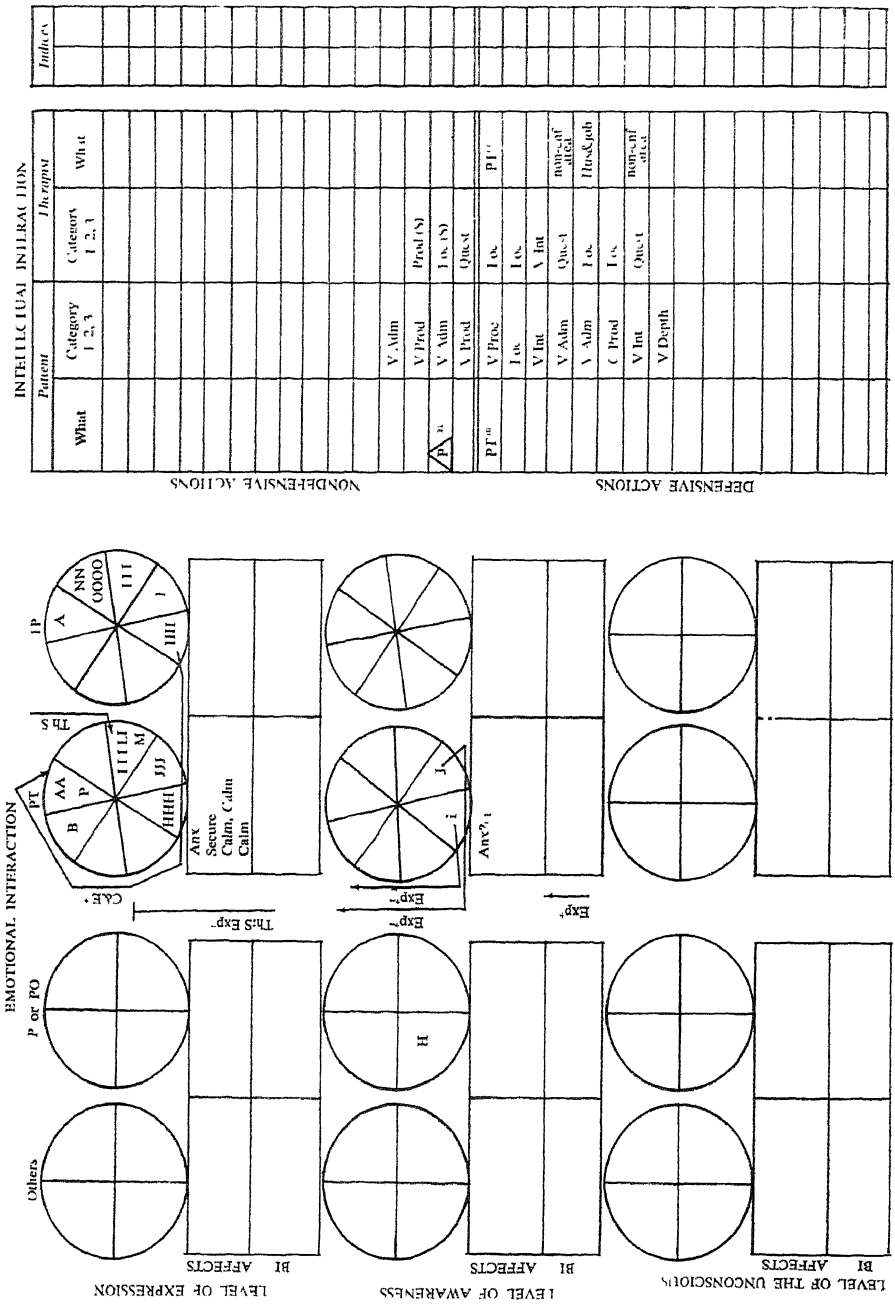
11. We have already indicated why we pool clinicians and why we do not offer figures for interclinician reliability

TABLE 14
CODES ASSIGNED TO STATEMENTS (275 TO 301) EXPRESSED IN
THIRD UNIT OF FIRST THERAPY SESSION

			Classification of Subject			Classification of Characteristic		
Speaker	Num-ber	Speaker Modifier	Subject	Category	Object	Score	Score Modifier	Category
Ther.	275	W	PT	A			V _{pt}	6
Patient	276	?	PT	A			VW _{th}	6
Ther.	277		TH	A			C	6
Ther.	278	W	PT	A			V _{pt}	6
Patient	279		We	A				7
"	280	NC	Miss Y	A			V _{Miss Y}	7
"	281		Hus	A				10-I
"	282		PO	A				10-I
"	283		PO	A				10-O
"	284	△	PO	A	Children	M*		5
"	285		PO	A	Parents	L*		5
"	286	C	We	A				7
"	287	C	We	A				7
"	288		Dr. X	A				7
"	289		Dr. X	A			C	7
"	290		We	A				10-O
"	291	C	We	A				9
"	292	WC	We	A				9
"	293		PO	A	Husband	N:Happy (w sex)		5
"	294	NC	PO	A	Husband	N:Happy (w sex)		5
"	295		We	A	We	N:Happy (w sex)		5
"	296		We	A	We	E*	<	5
"	297		We	A	We	N:E*		5
"	298		We	A	We	N:Sex		5
"	299		We	A	We	N:Sex		5
"	300		We	A	We	N:Sex		5
"	301	C	We	A	We	N:Happy		5

* Lettered scores from Interpersonal System of Personality (see table 4).

TABLE 15
SUMMARY OF STATEMENTS MADE IN FOUR FUNCTIONS ABOUT THE FIRST UNIT OF FIRST THERAPY SESSION



The patient and therapist interaction is tallied in the central sector (labeled PT and TP) and the patient's interactions with others on the left. Each level is divided into three variable systems; derived motives are tallied on the circles; affects are tallied in the upper box; and bodily motives are tallied in the lower box.

The positive, i.e., nondefensive, intellectual activities of the patient are scored above the double mid-line in columns labeled Categories 1, 2, 3. Defensive activities are listed below the mid-line. In the outside columns labeled "What" we tally the object of the variable, what it is about. The Fourth Category variables describing discharge or inhibition of impulse are drawn as arrows from the appropriate score in the motive sector. It must be remembered that the coding technician does not decide what is or is not defensive but simply classifies the evaluations of the clinicians.

EXPLANATION OF THE EXAMPLE

In Table 15 we observe that the clinicians have a lot to say about the non-bodily behavioral interaction between patient and therapist. The patient is seen principally as cooperative (L), docile (J), and modest (H) but also to some extent as dominating (A) and stubborn (B). The therapist is supportive (O, N), cooperative (L), and also modest (H). Less is said about affects and nothing about unconscious impulses although some expression of passive wishes (discouraged by the therapist—Express-Stimulate*) is mentioned. Only a single comment (H impulses in awareness) is made about the patient's relationships with others.

Both are considered to be operating defensively in the intellectual area. Eight defense scores (Admit-, Produce-, Internalize-, Depth-, etc.) are assigned to the patient. Seven inaccurate or inefficient

therapeutic interventions are mentioned for the therapist. The clinicians indicated that the therapist's questions about "nonconflictful material" and his focusing on the patient's husband and his job were contributing to the patient's defenses.

A different picture emerges from the tally sheet of the third unit (Table 16). The clinicians devote some attention to the patient's dissatisfaction with her husband but most of the comments still refer to the patient-therapist relationship. The therapist is more directive (A, P). The patient expresses more stubborn independence (B). In the intellectual interaction the therapist is quite active in focusing on and probing about the patient's frigidity. The patient is assigned five nondefensive actions but is predominantly defensive via intellectualization (Emotionalize-) about her frigidity.

INFERENCES WHICH MAY BE DRAWN FROM THE ILLUSTRATION

There has clearly been a marked change in both emotional and intellectual interactions between patient and therapist from the first to the third unit. Whereas in the first the therapist was principally supportive and the patient cooperative, in the third the therapist is more directive and the patient more stubborn. This change parallels one in the intellectual interaction. Whereas in the first unit the patient is seen as defensive mainly by denial and the therapist as failing to focus appropriately, in the third unit the therapist is actively probing about the patient's frigidity and she is talking about it but defending herself by intellectualizing. The therapist's intellectual probing is therefore described by the clinicians as directive and the patient responds by shifting from denial to intellectualization and also by becoming more stubborn. It is possible to calculate a number of summary ratios from the tally of the

clinical evaluation to quantify some of these relationships. These ratios can be either within the emotional interaction (e.g., how much does the clinician stress patient-therapist interaction as against interaction between patient and others) or the intellectual interaction (e.g., how does the patient's defensiveness compare with the therapist's) or between intellectual and emotional interactions (e.g., how much does the clinician stress one of these as against the other). Comparison of these indices indicates the amount and kind of change from unit to unit.

As an example, we will calculate here the percentage of defensive versus non-defensive intellectual actions (i.e., the number of codings below versus above the mid-line). For the first unit the patient's nondefensive percentage is 33: the third unit remains pretty much the same—31%. The therapist nondefensive percentage for the first unit is 30 but rises to 88 in the third unit. The clinicians clearly are more approving of the therapist's intellectual activity in the third unit, and the patient remains defensive, although she shifts from denial to intellectualization.

Tallying and Analyzing the Patient-Therapist Content

We now illustrate how this system of classification is used to quantify the recorded exchange between patient and therapist.

Table 17 presents the classification of the patient and therapist statements in the first ten-minute unit of an initial interview, i.e., the same interview as the one from which we have drawn our illustrations so far. In this table the patient's statements are indicated as "+," or in lower case (for the first eight categories); the therapist's statements are indicated as "x" or in capital letters. It is obvious from Table 17 that there are few clinical statements (12 for the pa-

tient, 2 for therapist) in this unit. The topics discussed were: miscellaneous aspects of the patient's behavior in therapy (22 statements in 10-M, i.e., Social miscellaneous of PT, i.e., patient in therapy) including details about her finding the office, patient's educational (33 statements in 10-I, i.e., Social intellectual-educational of PO, i.e., patient outside of therapy) and occupational (30 statements in 10-O, i.e., Social occupational of PO) characteristics, and occupational characteristics of certain social institutions (21 statements in 10-O of D, i.e., Sociological subjects).

Table 18 presents tallies of the number of statements made by the patient and the therapist in each of eight sectors of the classification matrix for the ten-minute unit just discussed as well as for the third unit from the same session (the same area for which we summarized the clinicians' evaluations). From these tallies it is possible to develop the index of relevance, which we described earlier as an indication of the probable importance for the development of insight, of the patient-therapist verbalizations for the statements made by each speaker during the ten-minute unit. We have already mentioned that a numerical indicator of relevance is obtained by assigning weights to each sector. The highest weight of five is assigned to statements which are most relevant to insight oriented psychotherapy—statements made by both patient and therapist about the insight oriented interventions and movements of psychic constituents from one level of our model to another. The lowest weight of zero is assigned to impersonal statements.

When the number of statements for each speaker in each area is multiplied by the weight and the weighted sum of all areas divided by the total number of statements by that speaker an index of "relevance" is obtained. An index of 5.0, for example, indicates that all statements were about dynamic processes of therapy.

An entirely impersonal discussion obtains an index of zero.

The indices for the first unit are: patient, 1.6, therapist, 1.54. This means that both participants are talking impersonally with practically the same "relevance." The manifest content of both patient and therapist during the third unit is much more personal and therefore "relevant" (patient = 2.24; therapist = 2.23).

Table 19 presents the values for these two units of some of the other indices which can be derived from a tally of patient and therapist statements, by making various subject and characteristic groupings and then comparing the tallies for the several groups either for patient or therapist alone or for patient and therapist with each other. We note that in both units the patient talks a good deal more than the therapist, but what

TABLE 17

TALLY SHEET FOR PATIENT-THERAPIST CONTENT OF FIRST UNIT OF FIRST THERAPY SESSION*

Th Totals	Pt Totals	Category	PT	Th	PO	KP We H	NKP	B	C	D	E
10	18	11 10-M	×××××××× ++++++++ +++ +		×		+++			× +	
1	3	G								× ++	
9	31	I			×××××××× ++++++++ ++++++++ ++	× +	× +			+++	
5	7	P				+	× +			×××× +++++	
14	39	S A R O	×		×××××××× ++++++++ ++++++++			+		××××× ++++++++ ++++++++	
		9									
1	8	7	×								
1	12	6	h		F, f, happy	d, i,					
		5	amazed		happy, happy	e, k					
					married,						
					bored						
		4									
		3									
		2									
		1									
41	110										
Th — Totals			10	0	17	0 1	2	0 0		11	0
Pt — Totals			17	0	54	1 5	5	1 0		27	0

Key: × = Therapist
Capital Letter = Therapist

+ = Patient
Lower case = Patient

this means we cannot say. In the first unit the index of clinical statements by the patient is low (11%) but it rises sharply in the third (56%). The same is true for the therapist (from 5% to 54%). This change in the index of clinical statements when seen in the light of the fact that the percent of personal subjects is high for both participants in both units (patient 74% and 84%; thera-

pist 73% and 75%) indicates that the rise in the index of relevance in the third unit is due not to a change from impersonal to personal subjects but rather that a shift to clinical characteristics has taken place.

They both talk some about the patient-therapist relationship in the first unit (patient 15% and therapist 24%) but the therapist talks much more about this

TABLE 18

SUMMARY OF STATEMENTS MADE BY PATIENT AND THERAPIST IN FIRST AND THIRD UNITS OF FIRST THERAPY SESSION

Non-Clinical Categories		Area 5 Index Weight = 2		Area 8 Index Weight = 0					
Clinical Categories	5 - 8	Area 2 Index Weight = 4	Area 4 Index Weight = 3	Area 6		Area 7			
	1 - 4	Area 1 Index Weight = 5	Area 3 Index Weight = 4	Index Weight = 2		Index Weight = 1			
		PT and TH	PO and Known Pers.	Not Known Persons		B	C	D	E
		A							

First Unit

Pt = 70 Th = 28		Pt = 28 Th = 11	
Pt = 2 Th = 1	Pt = 10 Th = 1		
A		B	C D E

Sum of Pt Statements = 110
Sum of Th Statements = 41
Therapist % = 27
Pt Index of Relevance = 1.61
Th Index of Relevance = 1.54

Third Unit

Pt = 26 Th = 3		Pt = 14 Th = 6	
Pt = 2 Th = 4	Pt = 46 Th = 9	Pt = 3	
A		B	C D E

Sum of Pt Statements = 91
Sum of Th Statements = 22
Therapist % = 20
Pt Index of Relevance = 2.24
Th Index of Relevance = 2.23

topic in the third unit than does the patient (21% versus 1%). Since the index of relevance has risen in the third unit and talk about patient-therapist relationship might be considered to suggest high relevance, the talk about patient-therapist relationship in first unit must not be about matters of importance. In the third unit the therapist may well be pushing for the discussion of more important aspects of the patient-therapist relationship, while the patient does not follow his lead. That the therapist index for derived motives for the third unit is 18% while the patient's is only 3% may corroborate this last idea.

Neither participant discusses bodily actions or symptoms in the first unit, but in the third unit the patient discusses bodily actions 21% to the therapist's 9%, while the therapist discusses symptoms 13% to the patient's 1%. It is possible that this means that the patient is talking about some bodily matter

which she does not label a symptom, while the therapist does so label the bodily action. Neither patient nor therapist, with a single exception, talk about categories 1 through 4 in either unit, so it is clear that they are not talking about the process of therapy, nor the patient's defenses, nor would one expect them to, since the therapy has just begun.

Comparison of Analyses of Clinical Evaluation and Patient-Therapist Content

We have repeatedly stressed that our principal goal is the comparison of clinical evaluation and patient-therapist content of the same segment of psychotherapy. Now we are in a position to make such a comparison for both the first and third units of this session as well as to compare the two units with each other.

We suggested in our introductory section that there are three kinds of comparison we wish to make: 1. To what

TABLE 19

INDICES CHARACTERIZING PATIENT-THERAPIST CONTENT DURING THE FIRST AND THIRD UNITS OF THE FIRST PSYCHOTHERAPY SESSION

<i>Definition of Index</i>	<i>Code</i>	<i>PATIENT</i>		<i>THERAPIST</i>	
		1st Unit	3rd Unit	1st Unit	3rd Unit
Index of relevance	Ir	1.61	2.24	1.54	2.23
% of total statements	Id%	73	80	27	20
% of Clinical statements*	Ps	11	56	05	54
% of personal subjects	Per%	74	84	73	75
% of statements about patient	Pt%	65	34	66	50
% of statements about patient-therapist relationship	Tr%	15	01	24	21
% of dynamic statements (categories 1 to 4)	Dy%	0	1	0	0
% of statements about derived motives	DM%	4	3	2	18
% of statements about affects	A%	4	22	0	18
% of statements about bodily motives	BM%	0	21	0	9
% of statements about symptoms	SY%	0	1	0	13

* This index and the following indices are obtained by dividing the total statements of the speaker in the appropriate subject and characteristic category by the total of the speaker's statements in all categories.

extent is the patient-therapist content covered by the clinical categories? 2. Insofar as patient-therapist content is so covered does it agree with clinical evaluation content?; and 3. Insofar as patient-therapist content is in nonclinical categories, do the two kinds of analysis of the raw data nevertheless illuminate each other?

We turn first to see how much of the two kinds of content fall into the clinical categories. The clinical evaluation content—as we have said we will almost always expect to be the case—is principally about the first four categories and the patient-therapist emotional interaction (fifth category). But patient-therapist content is not at all about the first four categories in either unit. The patient and therapist then are not talking at all about the inhibition or facilitation of insight. We have already said that, even in insight-oriented therapy we would not expect such talk early in treatment. Only later when patient and therapist do talk about such matters will we be able to see whether what they say corresponds with how the clinician sees their intellectual interaction.

And similarly while the clinicians talk a good deal about the emotional interaction between patient and therapist the two participants say nothing about it (their statements about patient-therapist relationship were not in the first eight categories).

In fact in the patient-therapist content there is relatively little of *any* kind of fifth category talk in the first unit, but a great deal more in the third unit. This shift is a result of the patient's talking about her husband and is paralleled by the fact that the clinicians also say much more about her relationship with her husband in the third unit than in the first. This change in content is of course reflected in the index of relevance which rises sharply for both patient and therapist from first to third unit.

Our second kind of comparison, the extent of agreement where both contents are in clinical categories, is possible only for the fifth category talk about her husband in the third unit. The clinicians say that she complains about her husband, is frigid with him, and has hostile feelings towards him. The tally sheet for the patient (not given here) indicates only that the patient has described her frigidity with her husband. It is clear that the clinicians have concluded from this that she is complaining about him and has hostile feelings towards him.

Our third kind of comparison is between the clinical evaluation and the patient-therapist content in nonclinical categories. Under this heading we can say that study of the clinicians' view of the emotional and intellectual interaction between patient and therapist makes clear what they regard as the reasons for the changes in patient-therapist content. The therapist is seen as strongly defensive in the first unit and as focusing on non-conflictual areas. He is friendly, even modest, and discourages the expression of feeling. He is seen, then, as participating in and contributing to the patient's defensiveness which is expressed by her talking in nonclinical categories. But in the third unit in which the patient-therapist content is more relevant the therapist is considered almost not at all defensive, more directive, and as focusing on the patient's frigidity. Although the patient responds by talking about her frigidity she is seen as continuing to be defensive by intellectualization and to show some stubbornness.

We can return now to our discussion of the problem of validity. We have already made clear that for these units we cannot compare what the clinicians say about the interaction between patient and therapist with what the two participants say since the protagonists do not discuss their interaction. Have we anything in the patient-therapist count which

serves as a measure of the clinical evaluation? The low relevance of the first unit seems to be a measure of defensiveness, or one might turn the proposition around and say that the judgment of defensiveness is borne out by the low relevance. But the third unit, judged equally defensive, though now the defense is intellectualization, shows a much higher relevance. Does anything in the count point to the intellectualization? Perhaps the discrepancy between the patient's talking more about bodily actions while the therapist talks more about derived actions can be understood to mean that the patient talks about bodily actions but in a detached way.

Only detailed study of various kinds of psychotherapy by our two counts will show us how to use them to gain some light on their validity.

FUTURE PLANS

This presentation of our system for the analysis of the process of psychotherapy has attempted to give an overview of our concepts and techniques. Further details of the collection, classification and analysis of data will be discussed in future publications.

Our aim was to develop a *method* for the analysis of the process of therapy. We now have such a method and we have offered data to show its reliability insofar as the coding of patient-therapist content is concerned. We are presently engaged in a study which we hope will teach us the ways in which our two counts can be usefully compared. We have obtained free clinical evaluations of each unit of a psychotherapy consisting of 17 interviews and we are coding clinical evaluations and the patient-therapist content of this therapy and studying the pattern of scores and indices as they vary from unit to unit and session to session.

The next step which the research calls for seems clearly the working out of a method of obtaining comprehensive clinical evaluations so that their reliability can be properly studied.

And then we plan to make one more study devoted to the method of analysis before applying the system to the study of specific problems in psychotherapy. We will compare our system of analysis with others developed to quantify the process of therapy. A number of methods have been proposed for such analysis, some relatively comprehensive, some dealing only with selected aspects of process. We plan to code and study by our method interviews which have been studied by these other methods; then we shall make blind predictions as to the results obtained by other methods and finally compare the analyses yielded by the several methods. In this way we hope to correct weaknesses and gaps in our system and to relate our methods to those of other researchers. We do not imply that a single system must be developed to be used in all process research. It is likely that a battery of methods will be developed and that it will be used variously, depending on the particular problem under investigation.

SUMMARY

We have argued that research into the process of psychotherapy requires study of the raw data both by clinical evaluation and by other less inferential methods so that each can be used in the validation of the other. Our own research goals were the development of a comprehensive set of the dimensions of psychotherapy and a quantitative measurement of the raw data of psychotherapy. Using as our point of departure a combined psychological-psychotherapeutic model which we derived in part from the study of free clinical evaluations, we presented the several groups of categories of sub-

jects and characteristics which form a comprehensive matrix for the classification of all statements in both clinical evaluation and patient-therapist content. These groups are the clinical categories, the special clinical categories, which two together make up the dimensions of the process of psychotherapy, and the non-clinical categories. We completed the presentation of our coding system by a description of the several groups of modifiers of statements. We proceeded then to a discussion of problems of units for clinical evaluation, coding of content, and summary analyses, and the issues of reliability and validity. Finally we illustrated the application of our system, demonstrating our working devices as well as concrete examples of the coding of the content of clinical evaluation and patient-therapist verbalization and the comparison of the two. We closed with a note on our future plans, stressing the importance of developing methods for the comparison of the two kinds of analysis of the raw data of psychotherapy and for obtaining comprehensive clinical evaluations so that proper studies of interclinician reliability can be made.

REFERENCES

1. Collier, R. M. Consciousness as a regulatory field: A theory of psychotherapy. *J. abnorm. soc. Psychol.*, 1957, 55, 275-282.
2. Dollard, J., & Mowrer, O. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, 42, 3-32.
3. Eldred, S., Hamburg, D., Inwood, E., Salzman, L., Meyersburg, H., & Goodrich, G. A procedure for the systematic analysis of psychotherapeutic interviews. *Psychiat.*, 1954, 17, 337-345.
4. Freud, S. *An outline of psychoanalysis*. New York: W. W. Norton, 1949, ch. 4.
5. Gill, M. The present state of psychoanalytic theory. *J. abnorm. soc. Psychol.*, 1959, 58, 1-8.
6. Gottschalk, L. A., & Hambidge, G., Jr. Verbal behavior analysis I. A systematic approach to the problem of quantifying psychologic processes. *J. proj Tech.*, 1955, 19, 387-409.
7. Jahoda, Marie, Deutsch, M., & Cook, S. *Research methods in social relations. Part One, Basic Processes*. New York: Dryden Press, 1951.
8. Leary, T. *Interpersonal Diagnosis of Personality*. New York: Ronald Press, 1957.
9. Mowrer, O. H. *Psychotherapy: theory and research*. New York: Ronald Press, 1953.
10. Pittenger, R. E., & Smith, H. L., Jr. A basis for some contributions of linguistics to psychiatry. *Psychiat.*, 1957, 20, 61-78.
11. Rapaport, D. The conceptual model of psychoanalysis. *J. Pers.*, 1951, 20, 56-81. Also in R. A. Knight & C. R. Friedman (Eds.), *Psychoanalytic psychiatry and psychology, clinical and theoretical papers*. Austen Riggs Center, Vol. I. New York: International Universities Press, 1954, 221-247.
12. Rapaport, D., & Gill, M. On metapsychology. *Int. J. Psychoanal.*, in press.
13. Strupp, H. A multidimensional system for analyzing psychotherapeutic techniques. *Psychiat.*, 1957, 20, 293-306.

A Tentative Scale for the Measurement of Process in Psychotherapy

CARL R. ROGERS, PH.D.

For the past several years I have been increasingly interested in the problem of finding concepts suitable to contain the phenomena of the *process* of therapy. I believe we have made progress in conceptualizing the outcomes of psychotherapy in ways which are specific, measurable, and rooted in a context of theory. This has led to promising research in outcomes (6, 7) which I am sure will be followed by further and more adequate research. But in regard to the *process* of psychotherapy we have had no satisfactory conceptions or theory. Studies of process have been largely studies of segmented outcomes, and these have not been too helpful in understanding what is going on in the fluid interchange of the interviews.

More than a year ago I determined to approach this problem of process in a more focused fashion. After exploring a number of avenues which seemed rather fruitless, I decided simply to become a naturalistic observer. Divesting myself of as many preconceptions as possible, I listened to many recorded therapeutic interviews, trying to listen freshly and naively to what was going on. The complexity was enormous, and at times I despaired of discovering any order in, or making any sense out of, the diversity of interaction, the multi-faceted flow of what was obviously a meaningful relationship.

Gradually my observations began to cluster, and I felt I could discern some order dimly glowing through them. I began to see, or to think that I saw, the nature of a rather fundamental continuum involved. Under pressure of time

for an address at the APA, I pulled together, with much hesitancy, the observations I had made of this continuum, and presented this as a very tentative continuous scale for the understanding of the flow of psychotherapy (4). Since that time I have been asking myself, Does this represent my experience? Does this way of conceptualizing the process have any operational meaning? Do I stand by its rather far-reaching implications? The current paper is an attempt to bring the reader up to date in this matter.

The Process Continuum of Personality Change.

The concept which impressed itself upon me as I listened and observed was a continuum which seems to apply to the whole spectrum of personality change and development, and not to psychotherapy alone. It is, very briefly, a continuum which reaches from rigidity and fixity of psychological functioning on the one hand, to psychological flow and changingness on the other. Let me try to spell this out a bit more clearly.

At one end of this tentative scale or continuum we find the individual living his life in terms of rigid personal constructs, based upon the ways he has construed experience in the past. He has little or no recognition of the ebb and flow of the feeling life within him, as it exists in the present. He is remote from his own immediate experiencing (an important term which I will discuss more fully). His communication, even in a receptive and acceptant climate, tends to be almost entirely about externals, and

almost never about self. The form of his communication tends to be: "The situation is . . . ;" "They are . . . ;" "They say . . ." If pressed he might say "My characteristics are . . . ," but he would almost never say "I feel . . . ," "I believe . . . ," "I am uncertain about . . ." He does not recognize himself as having problems. He does not perceive himself as a responsible agent in his world. He exhibits no desire to change, and on the contrary shows many signs of wishing to keep himself and his relationships to others and to his environment as unchanging and stereotyped as possible. He is characterized by stasis and fixity.

At the other extreme of this continuum we find the individual living *in* his feelings, knowingly, and with a basic trust in and acceptance of his feelings as a guide for his living. His experiencing is immediate, rich, and changing. His experiencing is used as a referent to which he can turn again and again for more meaning. The ways in which he construes his experience are continually changing in the light of further experiencing. He communicates himself freely, as a feeling, changing person. He lives responsibly and comfortably in a fluid relationship to others and to his environment. He is aware of himself, but not as an object. Rather it is a reflexive awareness, a subjective living in himself in motion. He has incorporated into his psychological life the quality of flow, of changingness. He lives fully in himself as an integrated, constantly changing flow of process.

Between these two extremes there lies a continuum which can be differentiated into any number of points. For purposes of illustration I endeavored to discriminate seven stages which I felt could be distinguished from one another. There is, however, no magic in this number, and one might equally well discriminate three stages, or fifteen, or even fifty if our observations were sufficiently refined.

The Basis of Scaling the Continuum

It is believed that this continuum of the process of personality change and development has a certain general usefulness simply as a concept. If however we wish to make of it an empirical scale in order to test various hypotheses in regard to process, then it is important to state the conditions under which behavior samples should be collected for this purpose.

It seems clear that individuals reveal themselves and their characteristics to differing degrees in different situations. It is therefore important that we endeavor to approximate some standard condition under which samplings of expressive behavior might be drawn. It is proposed that the standard psychological climate should be one in which the individual feels himself to be empathically understood, accepted, and received *as he is*. This also happens to be the situation which is hypothesized as facilitating the process in question, but for our present purpose its importance lies only in providing a climate or set of conditions which could be measured and equated for individuals at any stage on the continuum.

It is important to note the negative fact that the scale of which we shall speak does not apply to samples of expressive behavior taken from situations in which, at the time of sampling, the individual feels misunderstood, accepted only conditionally, or not fully received as he is.

The Nature of the Conceptual Model

As I have tried to fit the observed facts to some type of model, I have gradually recognized that they do not correspond well to the usual picture of a simple continuum. The analogy of a yardstick is not adequate. There are a number of separate elements or strands in this process of change which need to be taken

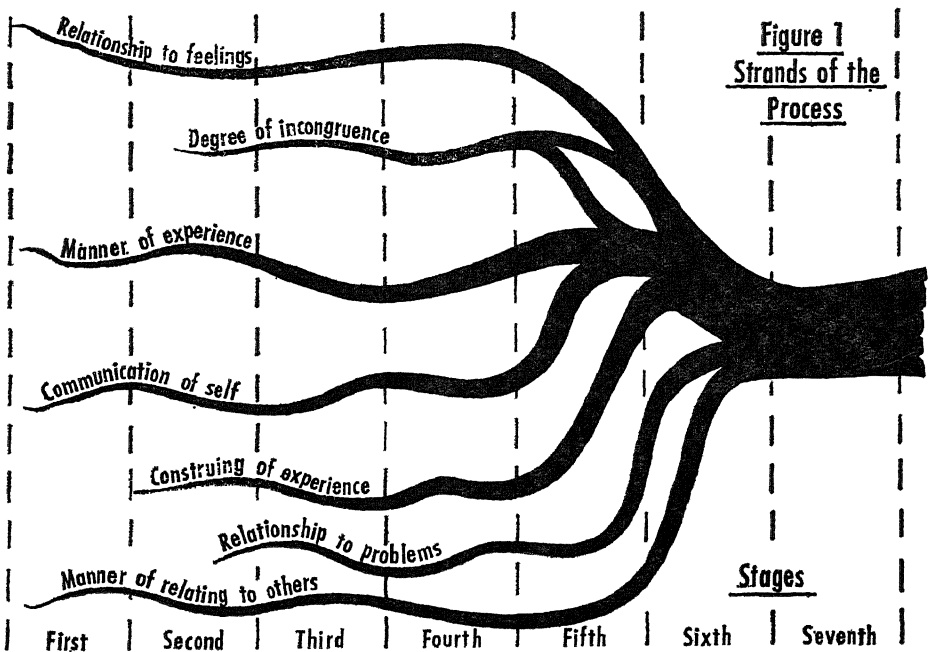
into account. But even these are not a series of yardsticks. The facts do not fit well the model of several parallel continua, on which different measurements might be summed or averaged to represent a stage on a more general continuum.

The distinctive point is that at the fixity end of the process, the various strands or elements are quite separable and distinct, and can be separately evaluated or rated. Whether the individual is exhibiting a rigid personal construct, or expressing himself on non-self topics, or describing feelings in a way which shows no direct ownership of them, these are rather clearly discriminable elements. But in the later stages of the process, the individual may be experiencing feelings with immediacy—knowing them and experiencing them being synonymous. These feelings are his expression of himself at that moment. They represent an immediately experienced change in a personal construct. Here all the previously separable strands are fused into

one moment, and to separate them is artificial.

It therefore appears that the most adequate diagrammatic model is of converging lines, separable at first, but becoming less and less separately distinguishable.

If I may push somewhat further this model of converging lines, I will use an analogy which in several ways may help to communicate both the quality and the form of the process I am trying to describe. Figure 1 pictures it. Here is a stream, originating in a number of completely separate sources in the foothills. If we think of these initial rivulets as being completely frozen, then we have the extreme of fixity in the process, the stasis end of the continuum. But if we think of the individual as being warmly received in his frozenness, then several trickles of flow and change begin. These may be frozen or dammed at some further point, but if the psychological climate continues to be favorable then



these separable rivulets increasingly flow into one another, becoming, at the optimal point of flow a unified stream of change in which the contribution of the separate tributaries can no longer be accurately distinguished, although all are present. Such an analogy appears to fit the nature of the process I am attempting to describe.

THE SIGNIFICANT STRANDS OF FLOW

The Relationship to Feelings and Personal Meanings

One of the discernible strands in the changing process is the relationship of the individual to the feelings and personal meanings which exist within himself. I have elsewhere defined a feeling as "an emotionally tinged experience with its personal meaning . . . a brief theme of experience, carrying with it the emotional coloring and the perceived meaning to the individual" (5).

At the rigid end of the process continuum the individual is largely unaware of his feeling life. Even in a receptive climate feelings are not described, and there is no evidence that they are in any way acceptable. Feelings may at times be *exhibited* in ways which seem quite obvious to the observer, but they are unrecognized as such by the individual. (Stage 1)

As we go up the scale we find feelings sometimes described as unowned past objects, external to self. (Stage 2)

Further on the continuum we find much description of feelings and personal meanings not now present. Even in describing these distant feelings, they are not apt to be pictured as acceptable, but tend to be seen as bad, unacceptable, or abnormal. At this stage when feelings are clearly *exhibited*, the individual may soon afterward recognize these as feelings. (Stage 3)

In the following stage we find feelings and personal meanings described as pres-

ent objects, owned by the self. There is considerable acceptance of these known, described feelings. Feelings of an intense sort are still described as not now present. Occasionally feelings are *expressed* in the present, but this occurs as though against the client's wishes. There is often a dim recognition that feelings previously denied to awareness may break through and be experienced in the present, but this seems to be a frightening possibility. (Stage 4)

In the next stage we find many feelings freely expressed in the moment of their occurrence and thus experienced in the immediate present. These feelings are owned or accepted. Feelings which have been previously denied now tend to "bubble through" into awareness, though there is fear and distrust of this occurrence. There is a beginning tendency to realize that experiencing a feeling provides a direct referent, to which the individual can turn for further meaning. (Stage 5)

The distinguishing mark of the next stage is the extent to which feelings which have previously been denied to awareness ("stuck" in their flow) are now experienced with immediacy and with acceptance. This experiencing is something which *is*, not something to be denied, feared, struggled against. In other respects this stage is similar to the preceding one, in that feelings are experienced and expressed in the immediate moment, with even greater freedom, and a deeper sense of ownership. (Stage 6)

In the final discernible stage new feelings are experienced with richness and immediacy, and this experiencing is used as a clear and definite referent from which further meanings may be drawn. Feelings are rarely denied to awareness, and then only temporarily. The individual is able both to live in his own feelings and personal meanings and to express them as an owned and accepted aspect of himself. (Stage 7)

Thus in this strand we find feelings and personal meanings at first unrecognized, unexpressed, though perhaps exhibited. They are then described as remote, unowned, and not now present. They are then described as present objects, with some sense of ownership. Next they are expressed as owned feelings, in terms closer to their experiencing. They are then experienced and expressed in the immediate present, with a decreasing fear of this process. At about this point even those feelings which have been previously denied to awareness bubble through, are experienced, and increasingly owned. Finally, living in the process of experiencing a continually changing flow of feeling becomes characteristic.

Manner of Experiencing

Another strand of the process, closely interwoven both with feelings and personal meanings, and with other strands to be described later, is the individual's manner of experiencing. This is a concept which has been elucidated by Gendlin.¹ He suggests that the term be em-

1. Gendlin's development of this concept of experiencing is one which I believe will bear significant fruit over the next decade or two. Not only is it a helpful concept in itself for purposes such as the present, but it represents another step toward what I believe will be the next trend in American psychology—a phenomenological, existential trend. This does not mean that I believe we will desert the logical positivist experimental tradition, nor the psychoanalytic dynamic tradition. It seems likely however that having selected from these streams of thought those elements with the most permanent value, we will use them to move on to new phases of discovery and assimilation in psychology, and it is my prediction that the place of subjectivity and the human encounter with life will receive more attention than at present. For this reason the discussion by Gendlin of the relationship of subjective experiencing to the logical positivism of psychology (1, Chap. VII) is refreshing indeed.

ployed to refer to the directly given felt datum which is implicitly meaningful. He says "Some initial sense of what the term 'experiencing' refers to can be communicated by calling it 'subjective experiencing.' It refers to an individual's feeling of *having* experience. It is a continuous stream of feelings with some few explicit contents. It is something given in the phenomenal field of every person." (1, Chap. VII). For example I can refer directly to the process in me which is involved in having a feeling, thinking a thought, deciding to act, etc. When I ask "what kind of an experiencing is this?" there is always an implicit answer, even though no explicit answer has as yet been conceptualized. Thus the answer might be "I am experiencing something vague and puzzling which I do not understand," or it might be a much more definite answer. Experiencing implies many possible conceptualizations. No one answer to the above question exhausts the possible conceptualizations.

In relation to our present interest in the process of change, we find that there is a great difference in the manner of experiencing at different stages of the process.

At the fixity end of the continuum immediacy of experiencing is completely absent. Conceptualizations as to the meaning of experience are all past formulations. The distance of the individual from his experiencing is very great. (Stage 1)

It is difficult sharply to differentiate the next stage but the individual is very remote from his experiencing, and reacts to internal and external situations as though they were past experiences, feeling them, rather than the present experience. Extreme intellectualization is one way of holding one's experiencing at arm's length. (Stage 2)

Perhaps the next development in this strand, is that the experiencing of situations is described as in the past. (Stage 3)

In the next stage there is an unwilling, fearful recognition that one is experiencing things—a vague realization that a disturbing type of inner referent does exist. Sometimes the individual recognizes an experience only shortly after the inner experiencing event. (Stage 4)

As we move up the continuum feelings are sometimes experienced with immediacy; that is, the individual conceptualizes and expresses his experiencing at the moment it occurs. This is a frightening and disturbing thing because it involves being in an unknown flow rather than in a clear structure. The only comforting aspect is that there is, in the fact of experiencing, a referent which can be symbolized and checked or rechecked for its further meanings and symbolizations. There is a strong desire for exactness in these conceptualizations. There may be a dim realization that living in terms of these solid referents would be possible. There also may be the realization that most experiencing occurs with some postponement after the event. (Stage 5)

Immediacy of experiencing, even of feelings previously denied, and an acceptance of being in process of experiencing, is characteristic of the next stage. The experiencing, in the immediate present, of feelings previously denied, is often vivid, dramatic, and releasing for the individual. There appear to be strong physiological concomitants. There is full acceptance now of experiencing as providing a clear and usable referent for getting at the implicit meanings of the individual's encounter with himself and with life. There is also the recognition that the self is now becoming this process of experiencing. (Stage 6)

In the final stage the individual lives comfortably in the changing flow of his experiencing. There is a trust in this process. The individual lives in terms of present experiencing, rather than interpreting the present in terms of the past. Differentiation between different experi-

enced referents is sharp and basic. (Stage 7)

Thus in this strand we find a continuum beginning with a fixed situation in which the individual is very remote from his experiencing, unable to draw upon or symbolize its implicit meanings. Experiencing must be safely in the past before meanings can be drawn from it, and the present is interpreted in terms of these past meanings. From this remoteness in relation to his experiencing the individual moves toward the recognition of experiencing as a troubling process going on within him. Experiencing gradually becomes a more accepted inner referent to which he can turn for increasingly accurate meanings. Finally he becomes able to live freely and acceptantly in a fluid process of experiencing, using it comfortably as the major referent for his living. In the furthest aspect of the continuum experiencing with immediacy is the major characteristic of the process of therapy. In such moments feeling and cognition interpenetrate, self is simply the reflexive awareness of the experiencing, volition is the natural following of the meaning of this flow of internal referents. The individual in this portion of therapy is a flowing process of accepted, integrated, experiencing.

The Degree of Incongruence

A third element which enters into the changing quality of the process of therapy is the change in the degree of incongruence. Incongruence is a concept which we have endeavored to define as the discrepancy which exists between what the individual is now experiencing and the representation of this in his awareness, or in his communication (5). Such discrepancy cannot be directly known to the individual himself, but may be observed. Its opposite is a congruence between the experiencing of the individual and the symbolization or conceptualization of this in his awareness.

In the beginning end of our continuum of process, we find a very considerable discrepancy between experiencing and awareness. This is observable to the trained diagnostician, or evident on projective tests. There is no awareness whatever of such discrepancy on the part of the client. (Stage 1)

A slight change in this picture is indicated when the client voices contradictory statements about himself as an object, with little or no awareness that these represent contradictions. (Stage 2)

Further up the continuum these contradictory statements are recognized as contradictory, and some dawning concern is felt about them (Stage 3)

Still further these contradictions are clearly realized and a definite concern about them is experienced. (Stage 4)

As another step in the process the contradictions are recognized as not simply diverse attitudes, but attitudes existing at different levels or in different aspects of the personality. Such phrases as "one part of me wants this, but another wants that," or "my mind tells me this is so, but *I* don't seem to believe it" indicate this kind of recognition of the nature of incongruence. (Stage 5)

In what appear to be significant moments of movement in therapy, there is a vivid experiencing of some aspect of incongruence as it disappears into congruence. That is, the individual is vividly aware of the inaccuracy with which he has symbolized his experiencing of some feeling, as he symbolizes it more accurately in the moment of full living of it. (Stage 6)

In the final stage incongruence is minimal and temporary as the individual is able to live more fully and acceptantly in the process of experiencing, and to symbolize and conceptualize the meanings which are implicitly in the immediate moment. (Stage 7)

The Communication of Self

Still another thread woven into this pattern of process involves the degree to which, and the manner in which, the individual communicates himself in a receptive climate.

At the frozen end of the continuum we find the individual unwilling to communicate self, even avoiding any expression which seems in any way revealing of self. Communication is about material entirely external to self. (Stage 1)

A bit further in the process expression begins to flow on topics which might seem related to the self, but which are handled as non-self material, e.g. "My education was good;" "my parents were insecure." (Stage 2)

The next step which can be differentiated involves a freer flow of expression about the self as an object, and about self-related experiences as objects. There may also be communication about the self as a reflected object, existing primarily in others. Past self-related feelings are described. (Stage 3)

In the next discernible step there is considerable communication of present self-related feelings. There is some expression of self-responsibility for problems. (Stage 4)

Further up the continuum we find the client freely expressing present self-related feelings. There is increasing acceptance or ownership of self-feelings, and a desire to be these feelings, "to be the real me." There is a clear acceptance of self-responsibility for problems. As the self is expressed as present feelings, there is less evidence of the self as an object. (Stage 5)

In the following stage the self exists in the experiencing of feelings. There is little awareness of self as an object. At any given moment the self *is* the experiencing. There is only a reflexive awareness. The self *is*, subjectively, in the existential moment. (Stage 6)

In the final stage the self is primarily a reflexive awareness of the process of experiencing. It is not a perceived object, but something confidently felt in process. It is not a structure to be defended, but a rich and changing awareness of the internal experiencing.

*The Manner in Which
Experience is Construed*

There are three other strands in the web of process which will be described much more briefly, since refined discrimination of points of the continuum does not yet seem possible in these elements. They will be described simply in terms of endpoints, with the implication that these elements exist in degrees, but that we are not yet able clearly to state these degrees.

The first is the manner in which experience is construed—borrowing Kelly's thinking (2) regarding personal constructs. At the fixed end of the continuum we find that personal constructs are extremely rigid, unrecognized as constructs, but thought of as external facts. Experience seems to *have* this meaning; the individual is quite unaware that he has construed experience as having this meaning.

In the process of therapy one can discern a gradual loosening of such constructs, a questioning of their validity, and an increasing discovery that experience has been construed as having such and such meaning rather than possessing this meaning inherently. Each such discovery naturally raises the question of the validity of such a construct.

In the moments of greatest movement in personality change, there is a dissolving of significant personal constructs in the vivid experiencing of feeling which runs counter to the construct. There is the realization that many personal constructs which have seemed to be solid guides are only ways of construing a moment of experiencing. The client often

feels "shaky" or "cut loose" as his solid foundations are recognized as constrictings taking place within himself.

In the flexible end of our continuum, experience is tentatively construed as having a certain meaning, but this meaning is always held loosely, and is checked and re-checked against further experiencing.

The Relationship to Problems

Another strand which may be briefly described is the way the individual relates to his problems. At the rigid end of the continuum, no problems are recognized and there is no desire to change. As there comes to be a recognition of problems, they are perceived as external to self, with no sense of responsibility for them.

As the process continues there is increasing recognition that the problems exist inside the individual rather than externally, and that the individual has contributed to their existence. Increasingly there is a sense of self-responsibility for problems. In the peak moments of therapy there is the living of a problem, the experiencing of it. It is no longer an object in itself. In the final phase the word "problem" is no longer particularly meaningful in the ongoing experiencing.

The Manner of Relating

Though the manner of relating to another is undoubtedly an important element of the process, it is not so easy to discern separable stages on this continuum. Suffice it to say that at the beginning stage of the process close relationships are perceived as dangerous, and the individual avoids them. During the process the individual becomes increasingly willing to risk relating to others on a feeling basis, and that in the final stages the individual openly and freely relates to the therapist and to others on the

basis of his immediate experiencing in the relationship.

THE USE OF THE PROCESS CONCEPT

Can the Scale be Made Operational? A Pilot Study

What has been presented thus far is an organization of observations, the beginning of a theory of the process of therapeutic change. Can these observations be formed into a reliable operational scale in order to test the hypotheses implicit in the theory? It is too soon to say, but I would like to mention two beginning efforts along this line.

With the help of Dr. Alan Walker the discursive observational account of the process of psychotherapy, contained in the initial paper, was translated into a more orderly schedule of stages of therapy, with the different strands to be considered at each stage.

At this point Mrs. Marsel Heisel planned and carried through a small pilot study. From a fully recorded case she selected five samples for judging. Her aim was to have the excerpts deal with the same major theme, yet be essentially random in their selection. Since this client had been much concerned with her relationship to her family, the excerpts were selected in the following way. The investigator took the first, seventh, thirteenth, nineteenth, and twenty-fifth interviews. Beginning in the middle of each of these interviews a search was made for the reference to the family situation nearest to this point. A sample unit consisted of a theme of family reference between two other topics, or if the theme was lengthy, cutting it off at some point of logical break, so that the samples would be too disparate in length. On the average the samples ran for 1.5 minutes of interview time.

These recorded samples (together with typed transcripts of them) were pre-

sented to eight judges, members of a seminar group. The excerpts were presented blind, using the method of paired comparisons. Each sample was listened to in relation to every other sample, and the task of the judge was to state which member of the pair he had just listened to was higher on the process scale. He was also asked to record the level of his confidence in making this judgment. After listening to the ten pairs he was also asked to rate the stage of the process at which each of the five samples, in his judgment, belonged.

The eight judges had a minimum of training for this task. Each had read the initial paper describing the concepts, and had participated in some general discussion on these topics. They had before them the form prepared by Walker to help visualize the characteristics of the different stages.

The range of the samples was not great, though clinical observation and judges' ratings indicated there had been movement. The five samples were rated, on the average, as extending from stage three to four (3.1 to 4.4).

In spite of the lack of training of the judges, and the restricted range of the samples, there was a significant degree of concordance among the judges. Kendall's w was .43, significant at approximately the .02 level. This seemed an encouraging indication that with some refinement of the scale and training of the judges, satisfactory reliability may be achieved.

Where the paired comparison difference between two samples was the least, so was the level of confidence in making the judgments. The level of confidence was greatest where the difference between the two samples was greatest. This too seems to indicate that there is a real and not a chance discrimination taking place.

It had been hypothesized that since the material was related to one major area

of concern there would be a positive correlation between the scale placement of the sample, and the interview from which the sample was drawn. In other words it was predicted that each successive sample would rate higher on the scale of process. Actually this relationship was positive, but not significant. This was largely because the sample from the nineteenth interview—a somewhat atypical sample—was judged lowest on the process scale.

It seemed reasonable to draw the tentative conclusion from this small pilot study that we are dealing with some real continuum, and that samples of client behavior taken in a receptive climate can be reliably scaled along this continuum.

A Beginning Validation of the Scale

Following this pilot study, Rablen and Walker, together with the writer, undertook a more crucial study with the scale of process, which is just being completed (8).

From earlier transcribed cases on which we had research data as to outcomes as well as counselor judgments, I selected six cases to represent a considerable range of outcome, with three cases representing marked progress in therapy and three minimal. I endeavored to select mostly brief cases, because at the time I thought we might be making ratings of all of the material in each case. The cases and the order in which I ranked them as to progress in therapy, using all the evidence available, was as follows.

1. Vib—9 interviews. Rated first in degree of movement or change in 1949 study (3) on objective evidence.
2. Oak—48 interviews. Showed marked objective progress in 1954 study (6).
3. Sar—4 interviews. Showed dramatic movement in four interviews.

4. Bebb—9 interviews. Showed objective progress in therapy, but decrement afterwards. Reported in 1954 study.
5. Sim—7 interviews. Ranked 7.5 out of 10 in progress in 1949 study.
6. Sketch—3 interviews. Ranked 9 out of 10 in progress in 1949 study.

From these cases single pages of the transcribed case were copied without any identifying information. These pages in the longer cases were taken from the second interview, the third interview, the third from the last, and the second from the last. In the three and four interview cases it was the first and second, and the next to last and last interviews which were sampled. In the three interview case two non-consecutive pages from near the end of the second interview were assigned at random to early and late conditions. In all but this last instance it was the next to the last complete page in the interview which was copied. This avoided "closing remarks" but sampled the presumably more significant half of the therapeutic hour.

This method of sampling gave us two pages from early interviews, and two pages from late interviews with each client, selected in such a way as to preclude bias. There were thus twenty-four pages, each identified only by a code number. These were placed in random order for presentation to the judges.

The two judges worked together for a number of hours training themselves on interview material from other cases in order to learn to make the discriminations called for in the scale. They tried rating several interview samples independently, to determine whether they had achieved inter-judge reliability.

They then turned to the randomly ordered twenty-four pages from these six cases. Working independently, with

no knowledge of the case or its degree of success or failure, and without knowing which four pages came from the same case, they performed a number of discriminations. The most crucial were these.

- a. They sorted the twenty-four pages into three equal groups representing lower, middle, and higher ratings on the process scale.
- b. They assigned to each page a rating as to its stage in the process of therapy, refined to the first decimal point. (Essentially a 70 point scale.)

The findings as to inter-judge reliability were as follows: When the twenty-four samples were sorted into low, medium, and high groups on the process scale there was 75% exact agreement, 25% one step disagreement, 0% two step disagreement. When the twenty-four stage ratings made by the two judges independently were correlated, the Pearson r was .83, significant at the .01 level. This is a very satisfactory degree of reliability.

A measure of validity could now be obtained in the following manner. For each case the mean rating on the process scale was calculated for both the early sample and the late sample. The mean change was then calculated, and the cases ranked from greatest to least change on the process scale. This ranking was then compared with the ranking of the cases on external criteria, given above.

When this was done a rho of .89 was found between the ranking based on rated movement on the process scale, and the ranking based upon external criteria. This figure is significant at the .02 level, though it must be interpreted with caution when the number of cases is so small.

Another way of considering the validity of the ratings is to compare the change on the process scale of the three cases selected as representing marked progress, with the change in the three

cases selected as representing minimal progress. In the first group the mean changes on the process scale were 2.3, 2.0, 1.5, a mean of 1.93. In the second group the changes were 1.15, .62, .30, a mean of .69. It will be noted that there is no overlap between the two groups.

From this small study it seems clear that satisfactory inter-judge reliability can be obtained in using the scale. The preliminary test of validity indicates that using very small samples of transcribed interview material, the scale differentiates satisfactorily between the degree of process movement in more successful and less successful cases. It seems reasonable to suspect that a higher reliability and even more satisfactory validity would be obtained if the ratings were to be based upon auditory samples of recorded interviews, rather than upon transcriptions.

One Implication

At times I feel very much sobered by the bold prediction which is implicit in the development of this conception of a scale of process in personality change. What has been presented hints at the possibility that a brief sample of an individual's expressive behavior, taken in a situation in which he feels fully received, can be analyzed to give us knowledge of where he stands on the continuum of psychotherapy or the even more general continuum of personality development and flow; and that this analysis may be possible without knowledge of the individual's genetic history, social milieu, personal background, personality type, psychological diagnosis, or length of time in therapy. The two small studies reported indicate that it is not unreasonable to expect that this implicit prediction may be fulfilled. To me this seems a startling development.

I should like to look at this for just a moment in perspective. It was only eighteen years ago that a group of us

at Ohio State gloated momentarily over having achieved our first aim—that of having obtained, for research purposes, a sound recording and a written transcription of a complete interview. Our pleasure was very quickly dimmed as we listened to, and read, the material. It seemed so complex, so formless, so fluid. How could it possibly be the basis of research? Was objectivity possible in this sphere? To me it seems that we have come a long ways from that day.

Conclusion

I have endeavored to give a simplified picture of a process continuum of psychotherapy and personality change. I have suggested the nature of the theory of process which is arising out of these observations. I have not tried to provide illustrations of this process as it is found in psychotherapy. Neither have I described the irregular rather than smooth nature of this process. Both of these points were covered in an earlier paper (4).

I have pointed out that on the basis of preliminary studies, it appears possible to give this theory of the process of therapy a reliable operational meaning which will enable us to test a variety of hypotheses as to the quality and nature of personality change as it occurs in psychotherapy.

REFERENCES

1. Gendlin, E. The function of experiencing in symbolization. Unpublished doctoral dissertation. Univer. of Chicago, 1958.
2. Kelly, G. A. *The psychology of personal constructs*. Vol. I. A theory of personality. New York: Norton, 1955.
3. Raskin, N. J. An analysis of six parallel studies of therapeutic process. *J. consult Psychol.*, 1949, 13, 206-220.
4. Rogers, C. R. A process conception of psychotherapy. *Amer. Psychologist*, 1958, 13, 142-149.
5. Rogers, C. R. A theory of therapy, personality, and interpersonal relationships as developed in the client-centered framework. To be published in volumes on the status of psychology in the U. S., being prepared by the APA. New York: McGraw Hill. (In press).
6. Rogers, C. R., & Dymond, R. (Eds) *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
7. Seeman, J., & Raskin, N. Research perspectives in client-centered therapy. Chap. 9 in Mowrer, O. H. (Ed.) *Psychotherapy: theory and research*. New York. Ronald Press, 1953, pp 205-234.
8. Walker, A. W., Rablen, R. A., & Rogers, C. R. Development of a scale to measure process changes in psychotherapy. Study in progress.

Discussion of Papers by Leary & Gill, and Rogers

DAVID SHAKOW, PH.D.

Let me start off rather inauspiciously with a remark which by this time is utterly commonplace but still worthwhile making—psychotherapy process research is outrageously complex. All of you who have done your homework, have wended your way through the elements, levels, orders, and moods of the Leary-Gill system and the strands and stages of the Rogers system—to which, not to discourage you, I expect later to add a *few* other aspects—are probably dismayed at the range of problems which Leary and Gill actually consider and I read Rogers as hinting at. I trust, however, that these two thoughtful papers have reassured you slightly that if the problem is complex it is not overwhelming.

For the short half-hour of my discussion I have the following broad plan: First, I should like to consider the assumptions which I believe to underlie present process research and then to outline the scope of process research in order to try to delineate some aspects of the field. Against this general background, I plan to discuss a few limited aspects of the two proposed systems. I hope to close with some remarks on the general nature of the attack on this area of psychotherapy research that appear appropriate at the present stage of development of the field.

Before going on to describe the scope of the problem I should like to set forth what most persons occupied in this field of research seem to me either explicitly, but largely implicitly, to accept. These agreements represent very considerable advances, made over a relatively short time and in the face of marked differences in theory and point of view. The

order in which I shall present the points does not at all reflect on their relative importance. I am sure that you can provide additions to the list which I shall give, but in general those I mention probably represent the most obvious points of agreement.

I suppose that if there is any categorical imperative in therapy research—at least in therapy process research—it is: “Love, cherish and respect the therapist—but for heaven’s sake don’t trust him!” Let me hasten to add that there is nothing invidious in this lack of trust. It is not distrust of the therapist either as therapist or as a human being that is here involved. The admonition falls rather into that universe of skepticism that scientists must have about themselves and others as investigators. It is the skepticism which recognizes the problems of achieving objectivity of report in any kind of situation, and the particular limitations of the human organism as a reporting instrument in interpersonal situations. Because of this fact, I find it difficult to accept the Leary-Gill statement about the possibility of studying psychotherapy process through therapist’s notes. Certain kinds of psychotherapy studies can of course be made by this method, but I doubt that such studies would fall into the compass of those we are considering at this session. I need not add that I am not raising questions about the value of such notes for accessory purposes.

The acceptance of such an imperative means that there is general agreement that the basic therapeutic operation must be recorded in some way. We owe a debt to Harold Lasswell (1) for first

suggesting the recording of therapy. We are even more indebted to Earl Zinn for his truly pioneer efforts in this direction over a quarter of a century ago. Carl Rogers and his groups at both Ohio State and Chicago deserve much gratitude because of their twenty-year-long persistent acceptance and practice of this notion. Their efforts have contributed more than anything else to make recording the generally accepted practice for psychotherapy research that it is.

While we are still on recording I might add a related second point about which there appears to be agreement. It is the notion that recording does not—when carried out under certain supportive and carefully arranged conditions—distort the data very markedly. At least the belief appears to be accepted that the recording does not distort the data to the extent of making us doubt whether we are any longer dealing fundamentally with the process we originally set out to study. Much could be said about this point but I shall be brief. We are here thrown into the midst of the effects on data of attitudes held towards the invasion of privacy. Our own studies (2) have indicated considerable differences in such attitudes, but generally in favor of accepting recording or the presence of an “outsider” when the context is professional as opposed to personal or social. Further, and this refers particularly to the therapist, the cultural climate makes a great deal of difference. There is a corollary to the Jones law—I mean the well-known one about keeping up with the Joneses—which says that if enough Joneses are taking up a new thing, then that new thing must be all right, despite underlying personal reluctances. Professional people are not immune to such influences. Those of you who are old enough professionally to have lived through the era since the Zinn recordings, have experienced the change in climate to which I refer. I see a similar

trend in relation to sound-film recording. Except for the expense of the process, I do not think that the period when this procedure will receive a reasonable degree of acceptance is far off. I do not, for instance, except in remote and isolated outposts, now find the look of horror on therapists’ faces which used to appear when such a proposal came up in discussion.

I think, however, that we should caution ourselves against accepting too blithely as a fact (the factual acceptance reciprocally influenced by the amount of recording we do) that recording does *not* introduce disturbance. This disturbance may be in either or both the patient and the therapist, and particularly for the latter in the case of intensive psychotherapy. Marked personality and individual differences exist in this sphere, differences which relate to the exposure tolerance of subjects, the length of time necessary for working through to total acceptance, and other aspects which you have undoubtedly considered for yourselves. The human psyche is a fearfully devious organ. In more cases than we think, unless we are aware of the problem and have an extended period for observation and checking, we are likely to miss the disturbing effects of the methods we use to obtain our data. Some brief anecdotal hints of such effects in the therapy field may be worthwhile mentioning. During one of our films we found a clear instance where the patient sought to draw the therapist into a kind of fellowship against “those people out there.” Apparently the patient used the experimenters, about whom he had of course been told, as a device for achieving closer relationship with the therapist. In another instance, the therapist, this time with a different patient, himself appeared to invite the patient into some such association against the “outsiders” apparently in reaction to the process of recording.

A third, perhaps more explicitly accepted, principle is that of objectivity. I believe we could get general agreement on the proposition that the principle of "intuition" in its more naive form no longer plays a role in therapy research. What has taken its place is an acceptance of the principle that many marginal cues are important forces in the interactions of the therapeutic process. All these cues are, however, now defined as potentially available for *public* examination. They are no longer limited, as they have been in the past, to being considered as intrapersonal communication within the person involved and not ever available to others.

The fourth point on which I believe we are coming to some agreement is the notion that *eventually* our data have to be as *full* as possible. I refer to the view that holds that sometimes we must get around to being able to examine all of what transpires in the therapeutic situation, all of that complex which *is* the therapy situation. What these "full" data are I shall discuss more *fully* later. Let me for the present leave the point as I have just stated it. Although we might achieve agreement on this goal as an ultimate one, we are of course far from accepting it in practice. In part, this lack of acceptance arises from purely practical considerations and as such is candidly faced by the investigator. He admits the partial nature of the data he has collected, but sees no other way of attacking the problem. However, some of the limited acceptance of this principle is, I believe, the result of what I consider to be a natural occupational defense of the psychologist (Whenever in this talk I say "psychologist" I, of course, use the term in the non-disciplinary sense—as referring to a person who is interested in the objective study of psychological phenomena.) We are so imbued with the complexity of the phenomena we have to study that if only

for purposes of survival we learn to simplify. This simplification is at times directed at reasonable goals—to hit directly at the *core* aspects of a problem. Some of the time, with great perspicacity, we are able to achieve this end. But more often I am afraid—and this holds particularly for psychologists in the narrower sense of the term—simplification becomes a goal in itself, the goal to obtain and maintain objectivity even when achieved at the cost of "coreness." Unconsciously we develop defenses against going after material that is too complex, and deliberately we carry out a process of data collecting which ends up with our only having easily manipulable data. Eventually this defeats our original purpose. Technical resources permitting, is not the important goal to collect data relevant to the problem no matter how complex, and then use our talent for simplification in the *analysis* rather than in the original selection of the data? How can we best deal with this apparent unconscious displacement of simplification at the analytic level to simplification at the collection level? In the context of our present meeting I am obviously raising the specific question as to whether we ought not to be collecting *more complete* naturalistic data in the therapy situation—data which could be studied repeatedly and at different times according to different hypotheses and for any one or more of its manifold aspects.

A fifth point about which there is rather general agreement relates to the diadic character of the therapeutic relationship—the importance of studying both the therapist and the patient. Here again, however, some defensiveness can be detected. I do not think we shall achieve real understanding of the therapeutic process until we study the therapist more fully. It is my impression that the great majority of us have been participants in a conspiracy to protect the therapist against being studied. And nat-

urally so—for where do our identifications lie? But with psychoanalysts increasingly becoming aware of and ready to talk about the problems of countertransference the atmosphere may change. I see the therapist's role studied best in investigations of the psychoanalytic process where the therapist provides after each session as much data as he can about his associations that were left unexpressed during the session.

I list the sixth point of agreement merely for completeness' sake since it has already been repeatedly mentioned. I refer to the problem of complexity. Is there anybody involved in process research who does not recognize the magnitude of the task?

The seventh point has to do with the place of theory in process research. I believe that there is increasing recognition of the importance of having some general personality or psychotherapeutic theory in order to make possible a comprehensive attack on the problems of process, although many researches can go along for a time on a purely empirical basis.

These then are some of the points on which I see a reasonable degree of agreement among all or most investigators in this area. Against such a background researchers go their different ways. It is to these differences that I wish now to come.

Many strategies of attack are being employed in process research. Some investigators emphasize the macroscopic, some the microscopic, some the comprehensive, some the partial. Theoretically, I suppose, one could at the macroscopic extreme take the view of the *whole* course of therapy as being one big blooming *gestalt*. Given the present limited intellectual abilities of even the best of us, it is difficult to see how objectivity can be attained with such an approach. At the other, the microscopic extreme, one could be concerned solely with the

detailed analysis of certain syllabic stresses through the course of the therapy. A range of somewhat smaller proportions I am happy to say—but still a considerable range—is represented in the group assembled at this meeting.

What kind of data does the therapy process actually provide? The full range ideally available to us at present would, I suppose, include the data from the actual therapy session in inter- and intrapersonal communication in four different modalities: (1) verbal contentual, (2) vocalization quality, (3) kinesic, and (4) covert physiological. These data are made available with different degrees of voluntariness, clearness and awareness. In addition to these data from the session we could also have available the post-session report and associations of the therapist about the session. Perhaps we might in addition even have the record of the therapist's periodic sessions with a consultant-supervisor. The importance of having all four modalities cannot be exaggerated since it is likely that these forms of expression are idiosyncratic both in relation to person and to stage of therapy. They are generally unlikely to be parallel, equivalent ways of communicating but rather inter-active, balancing, substitutive and complementary—parts of a *total* process of communication.

Given such an immense body of material there are of course many different ways of approaching it analytically. One way is along the comprehensive/partial continuum. The quite elaborate Leary-Gill method is only a selectively comprehensive attack, mainly directed at the contentual aspect of the verbal material. Others may choose to limit themselves to a more partial attack on the data—say to certain aspects of vocal qualifiers as they relate to changes of topic during the course of therapy.

Another way of approaching the task is along the micro-macroscopic con-

tinuum. Take, for example, the situation in sound-film. In the fifty-minute session there are approximately 72,000 frames. Dittmann's experience with the analysis of photographs made from successive third frames exposed in units of six, leads one to believe that at least in the affective sphere an inordinate number of frames may carry different kinds of information. If this is so, we get some notion of the enormous body of information which may be made available by the visual material of the hour alone. Much of it is probably in the form of cues which we are unable to report, or at least not able to report except under very special circumstances. Are we not compelled at least to hypothesize that such cues are an essential aspect of the therapeutic interaction?

Turning towards the macroscopic end of the continuum, one runs of course into the perennial problem of units both within modalities and across modalities. In the process of objectifying through part analysis of the data, what, for instance, gets lost by considering only the contentual material without the associated vocal qualifiers?

In relation to the macroscopic approach we encounter still another kind of problem, one that is rarely considered. I refer to the active as opposed to the passive attitude in data analysis. The problem stems from the notion that in the therapeutic interaction, communication occurs at a level which is interfered with by focused attention on details, the notion that what is called for is an attitude of "free-floating attention" (Freud) or of the "third ear" (Reik). If we accept this notion even in part, the question naturally arises as to how important such an attitude is also for the *data analyst's* understanding and analysis of the therapeutic process? One might argue that if the data analyst used an active attitude in his analysis of the data he might not be able to sense as readily

what was being communicated in a situation in which the therapist had practiced the kind of passive attitude we have described. In fact, would he not be dealing with a situation different in some respects from the one he was presumably studying? Is there some justification here for arguing that a data analyst *should*, for at least part of his data analysis, adopt the general attitude that the therapist takes during the therapy in the sense we have just been discussing, and that otherwise he runs the risk of losing important data on the therapeutic process?

Obviously this kind of analysis more appropriately belongs under the macroscopic approach. Microscopic work, by definition, requires focused attention. But even in relation to macroscopic analysis there is much doubt as to whether data analysts can work under such an attitude, at least without very considerable practice. I might describe an experience of my own in this respect. I was interested in the problem of what additional data might cumulatively become available to a data analyst from repeated viewing of a film. In order to get started I carried out a few limited explorations. I took two separate ten-minute blocks from a therapy session, viewed each film some dozen times successively and dictated a record of my reactions during the successive viewings. I carried the study out under two different self-instructions: One film excerpt was viewed under conditions of active analysis, the other film excerpt under conditions of passive analysis. For the latter I had a voice-key arrangement attached to an Audograph which permitted me in my "free-floating state" to record my reactions to the film while I was viewing it. I tried to make myself as comfortable as possible and to adopt as relaxed and as non-focused an attitude as I could. The reason for having an automatic voice-key arrangement was to reduce the ordinary manipulations of starting and stopping the machine which

would have tended to arouse me. The study was altogether too preliminary to report results. However, I did find that the passive attitude was an extremely difficult one to maintain at an optimal level. Among other difficulties, I would, for instance, tend to doze off and then over-compensate by more concentrated attention than the instructions called for. An experiment of this kind would have to be carried out carefully and with a variety of controls, but I was left with the feeling that it was worthwhile and perhaps necessary to do, despite the considerable investment in practice needed to achieve the proper state. It is quite possible that the *best* kind of a "Who listens?!" attitude may possibly provide us with more ready access to the marginal cues which presumably play an important role in therapy.

Some of you are naturally restive about my delay in getting to the two papers of this afternoon. And you have a right to feel so despite the fact that I have interpreted the discussant's task at a meeting such as this—where the papers are available beforehand—to be a more general one. Of course, one wants to say much about the papers. But what *can* one say about two such papers as those of Rogers, and Leary and Gill, except to offer some general remarks and a few spotty specific comments? These important papers do not lend themselves to short-term discussion. They each deserve the kind of interaction which comes only from workshop sessions with experts in the proposed methods—sessions in which there are exercises in applying the methods to actual therapy material. Only through the interchange provided by such a setting can one really discover the adequacies and inadequacies of the proposed methods.

Now let me consider the papers. Despite some similarities between them they are strikingly different approaches to process. For the purposes of the con-

ference we are fortunate to have been presented with such fundamentally different attacks on the problems of process. Rogers has insightfully emphasized and delineated some aspects of the relatively neglected area of intra-personal communication and the overall flow of therapy. He flows from peak to peak of the mountain range. Leary and Gill's meticulous stress on some of the innumerable features of progress along the trail gives us some notion of the complicated nature of the terrain that has to be traversed in getting from peak to peak. Both approaches are important. Because my own interest in psychotherapy is primarily in what it can contribute to the understanding of the complexities of personality interaction and change, I find the Leary-Gill approach more congenial both methodologically and theoretically. I also have the feeling that they are closer to the many complex phenomena of the psychotherapeutic process. But here temperamental differences among investigators must be recognized as givens and even encouraged.

Despite Rogers' own cautions to us about the oversimplified nature of his proposed model, I wonder whether the model in its present most complex form is not still much too simple. Just to react to one aspect of the model, I would ask whether the examination of a wider range of cases—including patients of the kind which are perhaps less likely to come to counselling centers—would not show some differences from a universal pattern of shift from rigidity to flow. Sometimes rigidity may actually be an advance over conditions of extreme looseness and be the path along which the patient has to go in order to achieve flexibility. Further, because the model attempts to deal with only part of the psyche, some consideration should perhaps be given to the concomitants of the "flow" state which have some of the outward appearances of "rigidities." I refer

here to the "automatizations" which become part of the psychic structure, the establishment of which psychotherapy has as one of its goals.

I am glad that Leary and Gill have described something of the history of their efforts, for in this way they have given us a notion of some of the complex problems that have to be faced by an investigator in this area. One cannot help but be impressed by their level-headedness, their meliorism, their readiness to shift their mode of attack when an attempted one did not work. I must, however, confess that if Dr. Gill—whom I know much better—had been presenting the paper, I might at least have whispered "Chicken!" to some of their shifts, especially to their shift in stand on the problem of units.

I am impressed with the procedure they finally adopted for making use of their clinicians. I wonder, however, whether the data provided to the clinicians might not in some way be made more similar to that provided to their technicians. On the one hand, in the present system the clinician presumably has less time with the data than does the technician. On the other, he has the advantage of the additional cues from the vocal qualifiers when he listens to the recordings. When recordings are not available and when the material is read by two staff members, it is difficult to tell whether an advantage or a disadvantage results. In any case the situation seems different in the two cases and would appear to interfere with the authors' being able to take fullest advantage of their unusual method.

I wish to conclude by coming out for an attack on the problems of psychotherapy process research through a wide variety of approaches—microscopic, macroscopic; comprehensive, partial; one modality, many modalities; one session, a full therapeutic series. This catholicity does not arise from a wish to emulate

Margaret Fuller in accepting the universe. Rather, I believe that this approach on a wide front is what we really need. As long as the investigator recognizes the limitations of his approach, deals as objectively as possible with his material, and makes it available to others in this form, then I think that at the present stage of our knowledge catholicity in approach is wisest. In a problem so vast, impressionistic views of the beast as well as microscopic reports on its earlobe blood vessels may both make potential contributions to the knowledge we need, if these meet the criteria we have set forth. I may be stating the issue a little extremely, but I believe that the fundamental point is sound. Those of you who hold to the advantages of greater partisanship might heed Emerson's—I believe—remark about Margaret. In our own relatively free society the attack on the problem appears to be along this wide front and you better had accept it.

I am convinced that a great many of the differences we find in approach to this complicated problem derive from personality differences among investigators. We do not recognize and accept this fact sufficiently. Just an example to bring out this point: I have been much interested in observing the reactions of persons professionally interested in therapy to descriptions of our sound-film project. One of the more common reactions revolves around concern for the large body of data that is collected. In some cases one even has the impression that some respondents—who are not at all interested in research and are therefore most unlikely to become personally involved in such a project—become rather panicky about the amount of data on a hypothetical and vicarious basis! It does not seem to occur to them or even to others, who are more research-oriented and who therefore might actually become involved in such research, that even though one has much material there

is no necessity for analyzing all the material immediately or even *ever* analyze all of it. The mere burden of the existence of such a body of material of this kind appears to be so great for some persons that one wonders sometimes whether it would not interfere with their analysis of *any* of it. Obviously such a person should select an approach which is limited and encompassable. He can make his contribution best in this way. Perhaps only liberated compulsives should be permitted to collect masses of material—relevant masses only, of course!

You have heard described the early stages of two quite different attempts to deal with the problem of process in

psychotherapy. I, myself, have merely tried to hint at the broader context of process research in which these efforts should be viewed. I have not touched upon the *major* contribution that process research could make, namely, to personality theory. I trust that sometime during our meeting we will get around to this topic.

REFERENCES

1. Lasswell, H. D. The problem of adequate personality records: A proposal. *Amer. J. Psychiat.*, 1928, 85, 1057-1066
2. Sternberg, R. S., Chapman, J., and Shakow, D. Psychotherapy research and the problem of intrusions on privacy. *Psychiat.*, 1958, 21, 195-203

Methods for Assessment of Change (A)

DR. GREENBLATT: May I comment on Dr. Rogers' discussion in terms of quoting from clinical experience—my clinical experience—and referring to two types of cases? One would be the schizophrenic case who frequently in the evolution of his schizophrenia passes through a compulsive rigid stage for a very, very long time, with final disorganization into a schizophrenic state. In his treatment he goes back through the same stages. In the schizophrenic state he may appear to be very free indeed in terms of being in touch with the most primitive kinds of feelings. When he moves during recovery to the compulsive state, he may appear clinically to be very tight, very defended and very rigid. That is one consideration.

A second would be the example of a case I have in mind, an obsessive compulsive lady who had three years of analysis and four years of analytically oriented psychotherapy. She is now just as obsessive-compulsive as ever. She is a hand washer. Yet to hear her talk you would think she had extraordinary freedom of communication with her emotional self. She seems to be able to speak out with insight and flow and feeling and all the rest of it, but she is not a whit better clinically.

The general question then would be something like this. Does your intuition about a conceptual picture of the whole process of psychotherapy apply primarily to cases that do not include this type of case, that might be less sick, or do you think it would be possible to bring these

kinds of cases into the framework of your formulations?

DR. ROGERS: That is a very good question and certainly time and more work with it will give us the answers. I have included in my thinking and in my listening a number of interviews with schizophrenic individuals. I don't know that I have any to match the second case you have mentioned. I noticed you said in schizophrenics you found them closely in touch with primitive feelings. I have found various descriptions which might make it sound as if that were so, but when you begin to use the kind of scale I am talking about, and ask whether the individual is experiencing and expressing and accepting his feelings of the moment, the experiences I have had and the recordings I have been able to listen to make me feel that is not so. Maybe I will discover I am mistaken. I certainly hold that very tentatively.

In much schizophrenic material the individual is actually quite remote from his experiencing. It is more that he appears to express primitive feelings, or he puts them in terms of others. Of course delusions and hallucinations are a good way of making experiencing very remote from one's self. So at least my present hypothesis is that this scale would cover that kind of individual. Whether it does or not will be an empirical question. At least I am not discouraged by the prospects thus far.

As to your comment on the obsessive-compulsive individual, I could not answer that without being able to listen to a recording or see a transcription of the remarks. It might well be a complete

Abridged. See Editors' note, page 49

exception to this scale. But I have been surprised that what seemed to me to be such a subtle concept has initially tested out rather well. I was quite prepared initially to try it as a scale and to see it be a complete flop. It is surprising that it has seemed to be usable.

DR. SEEMAN: This whole question of scaling has bothered me. I have often found that what looks like a linear scale turns out on closer examination to be much more complicated than a simple linear function. It may turn out that what you are talking about will be that way, too, although I think you are saying currently it looks linear.

To take an example that is fairly close to this, in our ratings of therapy with children, we did not have too much trouble in describing non-expressivity, that is, withdrawal. As soon as we moved out to expressive behavior, we got what looked very crudely like a Y shaped phenomenon, which on the one hand moves toward an expressive relaxed easy flowing sort of behavior which might be similar to one of your points on the scale, and on the other hand is a high degree of expressivity which looks like it gets shot forth in an expulsive way, with the tension released. One can almost see the accumulation of tension and the expulsion of that tension.

So we are left with this peculiar scale which starts out to be linear and then goes off in very different directions. From the point of view of a statistical analysis, it makes a neat question.

DR. SNYDER: I have a methodological question for Dr. Leary. I would like to ask you about the reliability and validity of your measures when you use ten-minute intervals. The reason I ask this is that in attempting to do a much simpler task—analyze, for instance, the effect on a simple continuum something like your own which would go from positive to negative—we have found that

groups of judges have tremendous difficulty if the time interval is more than half a minute, which is one twentieth of your ten minute session. So what kind of reliability and validity have you achieved in that long a time sample?

DR. LEARY: You must recognize that there are two aspects to your question. One has to do with the reliability of the technicians who are coding what the clinician has said about the ten-minute unit. The inter-technician agreement as to order, variable and level has been quite high (averaging ninety percent). The agreement is a function of how well trained the technician is, as well as a function of how well defined the orders and variables are. Inter-technician agreement has not been a problem.

The other aspect of your question has to do with agreement among clinicians in judging what is going on in the therapy session. One thing which tends to lower the inter-clinician agreement is that they talk about different aspects of the therapy process. One clinician may describe the emotional interaction. Another clinician may talk about the defensive processes against awareness. Another may speculate on the unconscious dynamics. Another might talk about the interventions of the therapist. They may all be observing accurately, but we obtain four different protocols with about four different aspects of the process. These are then scored by technicians and plotted on summaries and diagrams.

In order to improve clinician agreement, you have to (a) have clinicians who use the same clinical language, and (b) collect enough measures over a period time so that the tendency to attend to one dimension in one unit and a different one in the next balances out. So if anything we tend to get higher inter-clinician agreement when we use larger time units. That is, in one unit,

two clinicians may not agree as to the emotional interaction or one may mention it and one may not, but over the five units of the 50 minute session, they are more likely to agree as to what the pattern of emotional interaction has been.

Is that distinction clear or helpful at all?

DR. SNYDER: Yes. I wonder if you get involved in the question that you may have to train your raters, your technicians, so hard that they have learned an esoteric technique, but it is not replicable anywhere, and no one else could do it?

DR. LEARY: Again we have the two problems, getting the judgments from the clinicians and the coding ratings from the technicians. Your question, as I interpret it, has to do with the technician's classification of the clinician's content. We do not consider this a problem. I should point out that our technicians are not psychologists, are not necessarily college graduates. They are trained like mail clerks to recognize certain addresses (which are the sentences and phrases which the clinicians are sending them), then file them in the appropriate letter box. It has not proved to be an impossibly demanding task. We just trained a new technician, a high school graduate, in about two weeks so that her reliability is now up to the average of our technician group.

We have done very little as far as data collection is concerned. There are many challenging issues there. We may eventually train or limit our clinicians so that when they listen to 10 minutes or 15 minutes or 50 sessions, they will report back to us not in their own free language, but in the vocabulary and grammar of our system. Then we would be able to test clinician agreement, even clinicians of different schools, if they would drop their old school allegiances

for the time of their using our check lists or ratings. We could then test how closely they agree as to what is going on.

DR. WALLERSTEIN: I would like to ask a question of Dr. Shakow, and I suppose take exception to what he assumes we all agree on in regard to process research and maybe for argument's sake state an alternate assumption.

In talking about recording and the value of recording, his emphasis was that is where you get the fullest data. The built-in assumption is that the fullest data gives you the best view of process or can lead to those things which allow for the best view of process. Maybe we are working in another kind of project in which we are not getting the fullest kind of data (and maybe making up our own rationalizations in the sense that we ask what can be learned about process in other ways). We ask the therapist, in talking with him for three hours about a process that took 600 hours, "What strikes you most about this," and see what we get. What you get from him is the most central observation, and what for him has been the most pervasive aspect of the process. That is, what emerges is saliency in terms of what the person saw in it, and if you get another person, you get pervasiveness and saliency in terms of how another person saw it, and this we could postulate is also a view of process and a different view of process than that which emerges from the completeness side. So I don't know that we necessarily have to accept the assumption that only completeness allows a real view of process.

DR. SHAKOW: I don't think we actually disagree. I talked with Dr. Robbins in between sessions to make sure that we understood each other about what implications there were in the kind of research you were doing. I think actually we are talking about process in quite a different way. When I talk about

recorded data, we are interested in the analysis of the process, we are interested in the intimate details of the minute-by-minute changes which take place, the kind of thing that now is going along in the Rogers, but more in the Leary-Gill kind of analysis. I think what you people do is get the tops of the peaks, as it were. You cover a period of maybe four years of data. Obviously you are interested in the process, but process at a quite different level of abstraction and detail. I could not go to your data and from it get answers to the kinds of questions that I may be interested in, but you could come to my data and you could get answers to those questions.

DR. HUNT: Recording therapeutic process undoubtedly helps, but it does not bring the millennium, and clinical judgments are still often the measurements to be used. At the Student Counseling Bureau at the University of Illinois, our group has some 2500 hours of therapeutic counseling tape-recorded. It is still necessary, however, for someone to make judgments of what is recorded on those tapes. In a study we have just completed on reliability of judgments of therapist behavior, the judges contribute more variance than the counselor-client pairs, sometimes several times as much, and the judge-counselor-client interaction is considerable in about half of these 45 variables.

I find this a very disturbing finding. Perhaps our clinical judges are too inexperienced. They are graduate students. For this reason, we are repeating this study with the same batch of recordings, using four senior psychotherapists. It is my guess that the experienced clinicians will show the same disagreement that our students have shown, but this remains to be seen. In the meantime, this disconcerting finding argues that we must examine fairly carefully, even when we have our basic data from psychotherapy

tape-recorded, we must examine very carefully what aspects of it can be judged with reliability and what aspects can't.

DR. MATARAZZO: It seems to me what Dr. Hunt has said is in line with the findings of a recent study Dr. Raines published with Rohrer. The implications of their study were as follows. One could have a dozen psychiatrists interview a dozen patients. The dozen impressions of any *individual* patient given by these 12 psychiatrists would turn out to be all different. Furthermore, the results show that the psychiatrist, or psychotherapist or other type of clinician, tends in general to be redundant and to have his own biases, to see this or that other characteristic in *each* and every individual patient he sees.

I think in part this finding of Raines and Rohrer is very deceptively simple, and if all diagnostic-evaluative statements involve this pitfall I think all of us would have to give up. But I believe the statement that Dr. Leary made a few moments ago is important here. If the clinicians participating in this study will agree to talk in the same language and in the same frame of reference and use the same level of discourse: namely, forget their own biases for a minute, while looking at the same data, we conceivably might get much more inter-clinician agreement than otherwise seems possible.

DR. ROBBINS: I would like to continue from Dr. Matarazzo's point, because it disturbs me. We have had this experience in our own research. As I have tried to make eminently clear this morning, we all work from the same bias but our language was not the same. It took a lot of training. We try to be sure that our concepts of even certain commonplace words like transference were reasonably enough uniform, and we have agreed on what we call arbitrary definitions of these, and try to get accustomed to using these in a similar way.

When we do that, these variations have been significantly reduced. This is I think one of the essentials. We can all have a common point of view. We are Rogerian, Adlerian or what have you, but even within this there are very wide variations.

This has to do with the problems of the variability that comes out of the use of micro-units versus the macro-units in process observation. The smaller the units of observations I think the greater the variability. The larger the units of observation, the greater degree of perspective and uniformity, and you see that in any analytic discussion. I am speaking of psychoanalytic discussion, but I am sure it is true of other fields of reference. The other view has the trouble that you can't see the forest for the trees. While this has its positive values, it also presents problems.

DR. FRANK: I would like to ask for clarification on a very general issue which tends to slip by without adequate discussion. Dr. Shakow says we all agree that the processes of psychotherapy are very complex. I don't know where the complexity is. Every event in nature is complex. Agassiz had his students describe one dead fish for three days; there was no end to what they could see. A falling leaf is complex. The movements of the planets and stars in the Ptolemaic system are much more complex than in the Copernican one. Isn't the problem merely one of lacking guiding lines, not knowing what is relevant, and therefore trying to encompass everything? For example, let us take the hypothesis that all that really matters in therapy is that the therapist somehow arouses the patient's expectancy that he is going to get better. Then it might be very simple, because all these details might prove to be irrelevant. You would judge their significance in terms of whether or not they arouse this expectancy.

DR. SHAKOW: You are talking about psychotherapy in terms of outcome. When I am talking about processes of psychotherapy, I am talking about what psychotherapy contributes to the understanding of personality. I am interested in what happens when two people get together and what are the interactions. What, for instance, happens when something which was not remembered suddenly gets remembered, and you can follow its natural history and its effects on the person. You apparently in your present research interests are not interested in such a question. When I get involved in this kind of study, then I recognize that there are perhaps thousands of parallel processes going on during the process of a relatively short time of therapy. That is why I get impressed with the complexity.

You are quite right, the reason that the process of psychotherapy seems complex, is that this whole buzzing therapeutic world is such that we don't know how to tackle it. After we get some really fundamental principles perhaps we will be able to do something more. I don't know that we have real cues as to where to look for the understanding of personality from psychotherapy.

DR. SASLOW: I want to raise a question which belongs to both Dr. Shakow and Dr. Leary. I am not sure how far one wants to push one of the areas of agreement you mentioned, the necessity of a personality theory in studying the psychotherapy process. I raise that question as a way of asking Dr. Leary how much of psychoanalytic theory is really preserved in the system as you are now using it. You are using three categories: overt conscious behavior, expressivity with awareness and various kinds of unconscious behavior. Are you calling that Freud's psychoanalytic theory? It seems to me you are preserving just a very small amount of it, an amount to which

a great many people would not have the slightest hesitation in agreeing, whether they thought of themselves as psychoanalysts or not. But this is apart from the issue of the completeness of a personality theory which in Dr. Shakow's view is assumed to be necessary before you can do any work in this general field.

DR. SHAKOW: No. I think this should certainly be left up to the investigator to decide how extensive his theory is. I think what does happen is that very soon, sooner than if you start off just working empirically, one comes to some kind of theoretical system, either quite limited or quite broad. However, I would not insist that a theory is necessary.

DR. SASLOW: How detailed does such a theory really have to be, in this field, at the present stage of our knowledge, is more my question.

DR. LEARY: Yes, that is an interesting question. We have struggled with the issue of complexity and at times the system has been more complex than it is now. When we use this system as it is now defined, plotting or diagraming the five or six charts for a 50-minute session, or the 10 charts for two sessions, pretty soon we are snowed down with diagrams, and agreement among clinicians drops the more specific you get. So how to deal with this complexity? I think you should know first the limits of the complexity yielded by your measuring instruments and second how complex you have to be to test your hypotheses. For many studies you can retrench or compress, that is throw away some specificity. For example, in a couple of our studies all of our clinicians agreed that there was a fourth order activity of a defensive nature going on. That means they all agreed that the patient was verbalizing in a defensive way. But the specific variables they used to describe these fourth order defenses were different. Some said the patient was intellectualizing, some

said generalizing, some said denial, some said obsessive rumination and so forth. There was no agreement as to the specific terms. They all agreed that there was "fourth order minus" activity. We felt we could take all the fourth order variables and compress them into one scale, plus or minus for this fourth order, so we throw away the complexity. I think you have to know what complexity you are throwing away for any particular problem. In many studies we don't use half the orders. We use simple two dimensional grids which describe the events of that particular sequence of therapy.

DR. ROGERS: It simply occurs to me in listening to the confusion of this discussion—I don't mean confused people. I mean the welter of different points of view—that it would be a real contribution if we subsidized the production of a small number of cases with movies, tape recordings, pre- and post- batteries of tests and so on, and made those available on a research basis to the various members of the group who are working on different projects. Let us analyze those cases by the Leary-Gill kind of system or by the process scale I tentatively suggested. Let us try different measures of outcome. Let us use all of our research ingenuity, which obviously we are already expressing in diverse ways, on the same body of data, the same cluster of cases. I think it would help reduce the number of variables we were investigating because we would find out what you call "this" and I call "that" are essentially the same. Especially would this be true if these cases were selected on the basis of being in therapy with therapists of different orientations. They would help to throw light on the perennial question, "Do different therapeutic orientations produce different kinds of results? Or the same results by different means?" They would provide a body of

data in which we could test hypotheses from different theories, but on the same material.

It seems to me that such a step might be a very unifying element in the research in this field, and one that seems not impossible. It could be done at no more than the cost of some one of the projects that are now going on. So I toss that out for consideration.

DR. LUBORSKY: I would like to agree with that. I suppose there is a great deal of general agreement on how valuable it would be to have such cases studied in different treatment centers by different methods. I would like to hear your reply to a question derived from that: Whether you feel that the categories that you have devised apply especially to the type of treatment that the patients underwent, i.e., whether, because in the client-centered treatment major focus is upon experiencing of feelings, your categories were especially applicable to that focus and more so than would be true in other types of treatment.

I have another question that is somewhat related to that. You stress in your categories *how* the person experiences. You leave out *what* he experiences. You even make it a special virtue. I wonder why you say at the end we don't need to know much about any diagnostic formulation or his past history, and I wonder why you emphasize that. Is it for the sake of parsimony, or to say I can do it with "no hands" or something else?

Finally, why limit the period or unit of observation so much? A minute and a half, or a page are such short segments. Do you not drastically limit the understanding of the observer in that way?

DR. ROGERS: Let me answer your middle question first. Yes, it is purely from the point of view of parsimony. As Dr. Shakow points out, therapy is incredibly complex, yet our only hope rests

in finding parsimonious explanations of lawful events. That is why if we can find lawfulnesses which are relatively simple, then in my estimation that tends to put us ahead scientifically. That was my reason.

In response to your first question, whether the process scale I talked about is presumed to be applicable only to the kind of therapy with which I am most familiar, no, that is not my thought. My hypothesis would be that it would be applicable to any kind of therapy, but that could be put to empirical test. I could be quite wrong on that because there is no doubt that the basis for it is the soaking in of observations primarily from the kind of therapy with which I am most familiar.

One qualification that might be important in sampling interviews from other therapeutic orientations is, as you will recall I said, that the samples should be taken from periods when the person felt quite completely received as he is. I am sure there would be such periods in any form of therapy, but not every period would necessarily be of that sort in some therapeutic orientations.

Now your question as to why I deal with such very small samples of material, that was simply due to the fact that I wanted to give a very preliminary tryout to this scale to see whether it was worth pursuing further. No one can be more surprised than I to find that such minute samples seem to show a definite trend. I would not have been discouraged if nothing had come of the study, because I would immediately have said that our samples were too small to give a notion of what was going on. I would have carried the study further and enlarged the samples. I don't know quite what to make of the fact that such small samples proved to have meaning.

Again, one of the surprises to me is that such minute samples of the total

interaction in the case give anything like a reliable or valid indication of what is going on. My reaction to that is, praise be! That is going to save us a lot of slave labor, if relatively small samplings of therapy can give us objective clues as to the quality of the relationship or the quality of the process.

DR. ISAACS: I have been very interested in the question about how much material is necessary for study both in the sense of derived material and actual portions of a therapy session or a therapy series. My own attitude about what is microscopic and what is macroscopic has changed very radically over the last year and a half or so. When I first started out looking at sound movies, I thought perhaps we could look at a few sessions and get some data. It took a year's work on one session, on the gross postural pattern of the patient, to correlate that pattern with what could be inferred about his underlying attitudes. I thought perhaps this was microscopic.

During the last several weeks we have obtained a Moviola and can view movies frame by frame at slow speeds; we are down to 60 seconds which we have broken up into 8 units which we think are meaningful. I don't know how far this can go on. It is quite a question. I think it goes back to some of the things that have been said before about the value systems of the therapist and the value systems of the researchers.

DR. HUNT: I would like to go back to Dr. Roger's statement that he gets reliable measures of various therapist's attitudes from either lengthy or brief recordings. There is a conflict here. We have not been getting what I would call reliable measures of therapist behavior. We are using a modification of the Fiedler instrument for judging the therapist-client relationship. There are a few items on which our judges agree fairly well, but on at least half of the items we have

used, they do not agree well at all. At least *our* judges did not agree. Has Dr. Rogers used a different kind of variable? We have been dealing with such variables as the warmth of the counselor, and in fact this is the variable on which agreement among judges has been poorest.

DR. ROGERS: I will talk very subjectively on that. It seems to me that rarely in all of the research that different colleagues of mine have done over the years have we run into serious problems on inter-judge reliability. There have been some times when that has been a problem. But by and large inter-judge reliability has been very satisfactory.

I attributed this satisfactory experience to the fact that so many of our concepts are clinically based, so that we are asking for only a relatively low level of inference. To my way of thinking that is the way that science in this field is going to be advanced. I think when you get off into higher and higher levels of inference, then inter-judge reliability or interrater reliability drops like a shot. I question greatly whether we will be able to make progress along that line.

On the other hand, if you keep the concepts clinically based, and close to earth, then, as I say, I can only report our experience that we have rarely had difficulty in obtaining a satisfactory degree of inter-judge reliability.

DR. BUTLER: I would like to raise in a sense the issue raised by Dr. Greenblatt which seems to me to be a rather crucial issue. It strikes me this way. He talks about a kind of looseness of organization and behavior in an individual which occurs in the context of being pathological. Dr. Rogers was speaking about a healthy kind.

DR. ROGERS: A harmonious flow.

DR. BUTLER: Yes. You could put this in the context of semantics, but I think it is more important than semantics. I think there are two kinds of processes

of looseness which are perceptually distinguishable, but which we talk about the same way. That point I think is rather trivial. The one that strikes me as being fairly important is that this kind of talk guides our observations so that at one point we look at looseness and interpret it in any context as being pathological. I think a beautiful example of this in research is this concept of affective complexity as developed by the projective test people lately. It stemmed largely I think from some of our research at Chicago, where people have gone through therapy and been rated as being successful and been followed up and so forth and given the Rorschachs. When they were interpreted we found a looseness of organization which was interpreted as being pathological. Then when we confronted the Rorschach interpreters with the other criteria of success, they began to do some rather serious thinking about this, and at this point they developed new differentiations. I think Dr. Hedda Bolgar and other people who have worked on this now think they can differentiate between two kinds of loose or flowing organizations. So I suspect that somewhere, in Dr. Rogers' concept, is the notion of affective complexity which makes distinguishable what are probably the two different kinds of phenomena that Dr. Greenblatt and Dr. Rogers are talking about. This brings me finally around to the observation that perhaps

it is a good idea to do a lot of naturalistic observations and go back and listen to cases like this, people who are obsessive-compulsive and people who are freely expressive, and actually listen to a case like that.

DR. LEARY: I want to say something to Dr. Rogers' point with regard to complexity versus the clinical simplicity of ratings. I think there can be simplicity plus complexity if you are working with a multidimensional model. Only recently at the Kaiser Foundation have we worked with clinicians. We ducked clinical ratings for years because clinicians have complicated ways of thinking and talking and we can't get agreement and we could not (until the development of our system) analyze them. What we have done in the past was to obtain many simple measures which were sorted into the appropriate level. All our TAT analysis, for example, is accomplished by technicians who know nothing about psychology. Using, for example, the Interpersonal Check List, they make simple rating choices on interpersonal variables which are fed into level III. These are compared with what the patient says or somebody else has said. The patterning of relationships among variables, even when using simple measures, can give you a model or system which becomes very complex in the objectively determined discrepancies.

A Technique for Studying Changes in Interview Behavior^{1,2}

GEORGE SASLOW, M.D., AND JOSEPH D. MATARAZZO, PH.D.

A major problem confronting workers interested in establishing psychotherapy on a sounder scientific basis is that of the criterion of outcome. The state of affairs in this regard is little changed in 1958 compared to 1954 when one of us stated in a review of the literature:

Because of the unsatisfactoriness of our criteria for the outcome of psychotherapy, the present state of this subject is one of extreme confusion. We have to deal, on the one hand, with the enthusiastic conviction of psychotherapists and of many of their patients that the time, effort, and money expended have been associated with long-lasting fruitful changes in the life pattern of the patient; and, on the other hand, with the contradictory or disillusioning results of the efforts to study outcome in a systematic manner (36).

Only in the research program of Carl Rogers and collaborators (33) have significant relations between certain aspects of a defined psychotherapeutic procedure and criterion measures begun to emerge. In nearly all other studies, criteria of improvement which presumably have high face validity such as the following all seem to have failed as effective predictors (or concomitants) of changes purportedly associated with psychother-

apy: Rorschach and other projective techniques (*Annual Review of Psychology*, 1950-1958), job stability, marital stability, geographic mobility, and amount of health care (40), etc.

For these reasons research workers have continued to look for appropriate methods of assessing change in behavior. Many investigators have used the interview as their instrument of assessment. The advantages of the clinical interview are its obvious flexibility and uniqueness, so that every patient has an opportunity to manifest his own, presumably learned, interpersonal behavior patterns. Its major disadvantage as a research instrument is its notorious unreliability (1, 12, 19, 20, 31).

It has long seemed to us that some standardization of the interviewer's behavior in the clinical interview, combined with suitably precise recording of pre-defined variables, could enable one to surmount its major handicap of unreliability while preserving its dynamic nature (spontaneity, richness, multidimensionality, transference potentialities, etc.). Chapple (2-10) laid the groundwork for just such an approach in his description of the Interaction Chronograph method. A review of his work, and of the instrument itself, has been published by us (23).

The basis of Chapple's interaction method is an analysis of the time variable during the interview. After considerable work in the field, Chapple arrived at his conclusion that time was an important variable for describing human relations. He and his early collaborator, Arensberg, found that their field work as anthropologists was unduly hampered by the

1. Much of the material described in this paper has not previously been reported, while portions of several previously published studies have been included for continuity and clarity. We wish to acknowledge the participation of Samuel B. Guze, M.D., Ruth G. Matarazzo, Ph.D., Jeanne S. Phillips, Ph.D., and the assistance of Marilyn Ekdahl, M.A., Bernadene Allen, M.A., and Robert Taylor, M.A.

2. These investigations were supported by research grants (M-735, M-1107, and M-1938) from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

lack of precision and communicability of the various "subjective" variables which anthropologists (and other behavior scientists) were then using to describe human relations, in the family, tribe, interview-situation, etc. Chapple has described in every-day terms the importance of the time variable in human interactions.

When we had come to the conclusion that existing methods of appraising personality were inadequate, we decided to try a different approach. What was needed was an objective yardstick. We began, therefore, by agreeing to limit ourselves to those aspects of a person's behavior which could be directly observed and recorded. From an examination of our previous studies in evaluation of personality, we concluded that one measurable factor that seemed highly significant was time. The question then arose: what traits of personality express themselves in time (9, p. 199).

. . . the class of phenomena with which we are concerned comprises the (timing of) actions and interactions of individuals. It is now necessary to give a more precise definition of what this includes. As a matter of everyday observation, we see individuals coming together, and from the evidence of what we see and hear we unconsciously make certain judgments about their behavior. Such judgments are that one individual started to talk, and that the second individual to whom he was talking replied, and that both accompanied their speech with facial or bodily gestures (7, pp. 21-22).

We all know, as a matter of observation, that people have different rates (timing) of interaction. Some of our friends or acquaintances seem to talk and act very speedily as compared to ourselves; others are slow and deliberate. These characteristics of individuals are something we intuitively recognize, and we often are at variance with the rates at which others act. For example, where there are two persons in interaction, one whose actions are quick and speech voluble, and the other, slow and given to long, well-rounded periods, we are apt to find that the speedy one keeps interrupting the slow one, jumping in when the other pauses, and so on. If the slow individual is persistent, he may finally wear the other down, and our fast individual will subside into silence broken with a few "impatient" or "bored" remarks. Or conversely, the

speed at which the fast person acts may so upset the slow individual that it will throw him off his stride and he will later confess that he thought the other person "hard to talk to," "never stuck to the subject," "always interrupted" (7, pp. 31-32).

As we explained before, we do not infer and then attempt to record "feeling states" or "emotions" because we have no operations to deal with them and because we shall find that the quantitative analysis of interaction will in large measure describe such phenomena. In operations describing the timing of actions and events, we hope the reader will discover a useful and highly supple instrument. If the reader sharpens his powers of observation, he will see that in many cases people whom he does not like or cannot get along with, *say* exactly the same things that the people he does like say. So actors frequently take a short play, play it first as a tragedy and then, using the same words, play it as a comedy. Here the language is seen as unimportant, and the timing is the factor which makes the difference in its effect on the audience (7, p. 33).

From observation, we note how different people have different rates at which they originate (initiate) action. We all know cases of "bashful" people who will never speak unless spoken to. If spoken to they very frequently turn out to be very lively interactors. On the other hand, the "glad-hand artist," the "greeter," the man who speaks to everyone is a man with a very high origin (initiative) rate (7, p. 43).

The above quotations will suffice to give an introduction to Chapple's interaction theory of personality. He has taken the behavioristic position that personality can be assessed without recourse to intra-psychic and other currently popular psychodynamic formulations, and further that this assessment involves merely the process of observing the *time relations* in the interaction patterns of people. Accordingly, Chapple has indicated that this method, because of its objectivity, can lead to a *science* of personality. This view, it can be seen, is consistent with MacKinnon's conclusion that the most promising approach to personality assessment will come from a "field theory" which gives sufficient

weight both to "organismic" factors (the individual's behavior) and the "situational" or "field" (which involves the other interactees) variables (22, p. 43). Interestingly, Sarason (and others) have attempted to view the Rorschach test in a similar manner and have emphasized the examiner-subject relationship and the effects of this on the subject's productions (35).

NATURE OF THE INTERACTION
CHRONOGRAPH FOR MAKING
INTERVIEW OBSERVATIONS

Essentially the Interaction Chronograph is a device which allows an observer to record, in time units with a high degree of precision, the behavioral interaction³ of two individuals. The 14 variables, definitions of which are given in Table 1, are objectively recorded by an observer who activates a series of electrically controlled counters which are connected to two keys, one for the interviewer, the other for the subject. Each key is depressed by the observer whenever the designated individual is talking, nodding, gesturing, or in other ways communicating (interacting) with the second person. Values for these variables are cumulative and can be abstracted from the printed record of the total interview with little difficulty. Some of these variables may seem unusually arbitrary, since they represent algebraic sums of two variables rather than individual measures of each of these variables. Chapple, in developing his interaction theory of personality, considered these derived variables more useful than the first-order variables from which they were obtained. In addition to containing individual

3. Records include only such behavior as number of utterances, number of interruptions, their durations, etc., and not the content of the verbalizations.

TABLE 1

DEFINITIONS OF THE INTERACTION
VARIABLES STUDIED

1. *Pt.'s Units*: The number of times the patient acted.
2. *Pt.'s Action[†]*: The average duration of the patient's actions.
3. *Pt.'s Silence[†]*: The average duration of the patient's silences.
4. *Pt.'s Tempo[†]*: The average duration of each action plus its following inaction as a single measure.
5. *Pt.'s Activity[†]*: The average duration of each action minus its following inaction, as a single measure.
6. *Pt.'s Adjustment[†]*: The durations of the patient's interruptions minus the durations of his failures to respond, divided by Pt.'s Units.
7. *Interviewer's Adjustment[†]*: The durations of the interviewer's interruptions minus the durations of his failures to respond, divided by Pt.'s Units.
8. *Pt.'s Initiative*: The percent of times, out of the available number of opportunities (usually 12) in Period 2, in which the patient acted again (within a 15-second limit) following his own last action.
9. *Pt.'s Dominance*: The number of times (out of 12) in Period 4 that the patient "talked down" the interviewer minus the number of times the interviewer "talked down" the patient, divided by the number of Pt.'s Units in the Period.
10. *Pt.'s Synchronization*: The number of times the patient either interrupted or failed to respond to the interviewer, divided by the number of Pt.'s Units.
11. *Interviewer's Units*: The number of times the interviewer acted.
12. *Pt.'s Quickness[†]*: The average length of time in Period 2 that the patient waited before taking the initiative following his own last action.
13. *No. of Interruptions*: The number of times one interactee interrupted the other during the total interview (or a period thereof).
14. *Length of Interview*: The duration of the interview in minutes.

[†] Values for these variables are recorded in hundredths of a minute. To convert to seconds multiply the given value by 0.6.

counters for each variable, the Interaction Chronograph has a "signal" counter which, when pressed by the observer after a pre-arranged signal from the interviewer, functions as a marker to record the start of different periods of the interview.

THE INFLUENCE OF INTERACTEES UPON EACH OTHER DURING INTERVIEWS

In his early study of interaction patterns during interviews, Chapple placed little restriction upon the interviewer other than that he should use a non-directive or counseling interview of the type described by Rogers (32). Chapple soon discovered, however, that every interviewer was different not only in the way he behaved but in the results that he obtained from the same subject (4, p. 296; 5, p. 23; 9, p. 203). Thus it became clear not only that different interviewers have different interaction patterns when behaving in their own characteristic manner but that, as a result of these interviewer differences, different interaction patterns were elicited from the same patient when seen by two different interviewers. This was apparently true even though one could perceive no difference in what the interviewers said as judged by stenographic transcripts and even though they were following a supposedly uniform style of interviewing. Note the similarity between this and the recent finding by Raines and Rohrer (31) that differences in impressions of the *same* patient by different psychiatrists are due to personality differences in the psychiatrists, themselves.

Chapple's evidence for this subtle, but theoretically very important point, came from two reliability studies he conducted (4, pp. 300-301). The first was on 40 men who were interviewed (individually and independently) by two psychiatrists, each of whom used his own interview pattern; and the second series was on 12

persons interviewed under similar conditions except that the two interviewers behaved in a standardized manner (to be described below). The Pearson reliability coefficients for the Tempo and Activity scores went from .53 and .79, respectively, in the first study to .93 and .95 in the second. Thus, the effect of standardizing certain features of the interviewer's behavior which hitherto had not been dealt with explicitly was to elicit almost identical patterns (for these two variables) from individuals interacting with two different interviewers.

An equally striking demonstration of the possible interviewer-effect on the conversation pattern of the second person becomes evident if one re-analyzes some data Chapple presented in 1940 (7, p. 35). Chapple gives mean values for the Action variable during conversation for individuals A and B who met together eight times over a period of two months. The range of A's daily average duration of Action during these eight conversations was from 3.10 to 7.81 hundredths of a minute, while B's Actions ranged from 2.89 to 4.33. If one computes a rank order correlation from the data Chapple presents in his table, i.e., the eight daily values for A's Action as against the eight paired values for B, one finds a striking value for rho of .83. Thus A and B were having a subtle but marked effect on each other's duration of utterance. This finding has considerable implication for planned behavior therapy, i.e., an interviewer might, by plan, increase his own average Action with low activity patients and decrease it with patients of abnormally high activity, and maintain this "prescribed" behavior therapy over a planned number of months.

That experienced therapists can be flexible in their interaction patterns with different subjects and thus, theoretically at least, practice prescribed behavior therapy was demonstrated in a study by

Saslow, Goodrich, and Stein (37). These investigators studied, with an old form of the Interaction Chronograph, the behavior of a single therapist with 12 psychiatric patients he was interviewing, and in 7 conversations with three of his colleagues. The range of interaction behaviors available to this therapist, as well as the significant differences in his behavior with the two groups, are well illustrated.

Goldman-Eisler, working at Maudsley Hospital in London with an ink-writing version of Chapple's Interaction Chronograph, has added further evidence for the influence of interviewers' behavior on the behavior of interviewees. In her first study (13), using seven members of the Psychology Department as subjects, she demonstrated that, in free conversation, each had interaction patterns (short and long silences and short and long actions) which were rather stable for any particular individual. She also found that the silence variable (this differs a little from Chapple's) was a somewhat more stable personality characteristic than was action. Her next study was designed to investigate individual differences between interviewers (three senior psychiatrists) and the effects of these differences on the interaction patterns of patient-interviewees (14). Each of her interviewers used his own pattern of interviewing, i.e., the latter was non-standardized and was one of the variables under study. She confirmed Chapple's findings in that the three psychiatrists each had his own individual interaction pattern regardless of the type of patient he was interviewing (depressed versus active patients), although this pattern could be adjusted to the type of patient (within the limits of each interviewer's own pattern) somewhat. Furthermore, her results are striking in their demonstration that these three doctors influenced the interaction patterns (total activity, and ratio of short silences to long silences) of the same 10

patients in different ways. Thus depressed patients talked more with one doctor than with another, while these same doctors had opposite effects on talkative (anxious) patients. Goldman-Eisler notes, "Thus it would seem from the above that the depressed patient responds best to active stimulation, and the one who talks easily gets the best chance with the passive interview technique" (14, p. 670). Here, again, is evidence suggesting that planned "behavior therapy" may ultimately be possible.

The similarities between these various results obtained in complex, molar (interview) situations and the laboratory findings on experimenter-produced modifications in verbal *content* of Greenspoon (17), Salzinger and Pisoni (34) and Sidowski (41) are very striking. Thus Sidowski, by the simple procedure of blinking a light as a reinforcer each time his subjects produced a plural word in a situation in which they were instructed only to "start saying words," was able to demonstrate the effect of learning "without awareness" on the part of the learner. Greenspoon obtained similar results by using the verbal response "umm-hmm" by the experimenter each time the subject gave a plural response. Salzinger and Pisoni, by use of such verbal reinforcers by the interviewer as "mmm-hmm," "uh-ha," or "I see" each time an affect response was made by the subject were able to produce a significant increase in this class of response in a group of schizophrenic patients. For an excellent review of these and some several dozen similar studies, the reader is referred to the recent paper by Krasner (21). All of these results are provocative and considerably more research will no doubt be done in this area.

Goldman-Eisler indicated that her Interaction Chronograph personality descriptions of the three psychiatrists possessed some validity since the rank order

correlations between objective values so derived and opinions obtained *independently* from seven psychiatrist colleagues "were in complete conformity" (14, p. 670). Considerably more research of this type is indicated, however. Goldman-Eisler next reported a study in which she demonstrated individual differences among such language (content) variables as word rate, self-reference terms, verb-adjective ratios, etc., among the three psychiatrists, as well as within the patient group (15). Her latest study reports differences in speed of talking among individuals (normals and patients) and also the sensitivity of this variable to the interaction characteristic (rate of utterance) of the other person (16). Thus her earlier results were confirmed for another interaction variable.

This series of studies seems to confirm MacKinnon's conclusions of "specificity and generality of behavior" (22, p. 43); i.e., that a subject's behavior is a function not only of certain fixed organismic variables, but also is dependent upon the prevailing stimulus conditions. Thus these studies help to define the parameters of this specificity (situation) and generality (organismic variables).

NEED FOR STANDARDIZATION OF THE INTERVIEW SITUATION FOR RESEARCH PURPOSES

Chapple's own experiences with interviewers, each of whom used their own individual interviewing style, and the results of the laboratory experiments reviewed by Krasner, all indicate the need for some control or standardization of the interviewer's behavior if the clinical interview is to be used as a research tool. To this end, Chapple suggested some rules to guide the interviewer's behavior. In addition, by prescribing that the interviewer behave in a number of different specified ways as the interview proceeds, it was possible to sample a larger portion

of the interviewee's repertoire of responses. Therefore, the standardized interview is divided into five periods, with Periods 1, 3, and 5 as free give-and-take periods, and Period 2 (silence) and 4 (interruption) as *stress* phases of the interview (5). The characteristics of the standardized interview are shown in Table 2, whereas the "rules" governing the interviewer's behavior have been standardized by us (23) and are given in Table 3. As can be seen, with its two stress periods and three non-stress periods Chapple's standardized interview provides a means for eliciting a sample of patients' behaviors (dependent variables) in a complex but miniature, molar, interpersonal situation, the characteristics of which to a certain degree are objective and pre-defined (independent variables).

TABLE 2
CHARACTERISTICS OF THE STANDARDIZED
INTERVIEW

<i>Period</i>	<i>Type of Inter- viewing</i>	<i>Duration of Period Fixed Duration</i>	<i>Variable Duration</i>
1	Free	10 min- utes	
2	Stress (silence)		12 failures to respond, or 15 minutes, which- ever is shorter
3	Free	5 min- utes	
4	Stress (inter- ruption)		12 interrup- tions, or 15 minutes, which- ever is shorter
5	Free	5 min- utes	
Total		20 min- utes	plus a maximum of 30 more min- utes

RELIABILITY OF THE METHOD OF THE PARTIALLY STANDARDIZED INTERVIEW

Before the usefulness of such a research instrument can be determined, its reliability must be known. The complexities of a psychiatric interview make the reliability problem one of multiple aspects, each of which needs independent study.

Observer Reliability. First to be mentioned is the reliability of the observer's recording of the interview. He is an integral part of the instrument and the final results will be no more reliable than the reliability or accuracy of his input.

The availability of two Interaction Chronographs in the personnel department of a large department store made

TABLE 3

RULES FOR INTERVIEWER

Periods 1 to 5 (all periods):

- a. Interviewer introduces each period by a 5-second utterance (following his signal to the observer).
- b. All interviewing must be *nondirective*. No direct questions, no probing or depth interviewing. Interviewer can reflect, ask for clarification, ask for more information, introduce a new topic area, etc. In general, interviewer's comments should be non-challenging and open-ended and related to the patient's past comments or to some new, general topic.
- c. All interactions must be verbal only, or verbal and gestural at the same time; i.e., interviewer cannot use head nods and other gestures alone. This rule simplifies the observer's task.
- d. All of interviewer's utterances must be of approximately 5-seconds' duration.
- e. After patient finishes a comment or other interaction, interviewer *must* respond in less than 1 second, except as otherwise noted in Period 2.
- f. Each time patient interrupts interviewer, the latter must continue to talk for 2 more seconds. This rule insures more explicit definition of a patient's ascendance-submission pattern than would be possible if interviewer "submitted" immediately.

Periods 1, 3, and 5:

- a. Interviewer must never interrupt patient.
- b. If after interviewer makes a comment patient does not respond, interviewer must wait 15 seconds and then speak again for 5 seconds.

Period 2 only:

- a. Interviewer must "fail to respond" to last interaction of patient a total of 12 times (or for 15 minutes, whichever is shorter).
- b. After interviewer has been silent for 15 seconds (and patient has not taken initiative) interviewer makes another 5-second comment.

Period 4 only:

- a. Each time patient acts, interviewer must interrupt patient for 5 seconds for a total of 12 times.
- b. Interviewer's interruption should begin about 3 seconds after patient has begun his interaction.
- c. After having interrupted patient, if the patient continues through the interruption (does not submit), interviewer will not interrupt again until patient has finished his utterance, i.e., interviewer will interrupt patient only once during each utterance of the latter if patient does not "yield."
- d. The Period is ended after 12 interruptions or 15 minutes of attempting to obtain these.

possible simultaneous but independent recordings of the same standardized interview by two observers (O's). One of the O's had had approximately two years of experience (involving many hundreds of interviews) observing the standardized interview in an employment setting. The second O was relatively inexperienced, having recorded only some 10 practice interviews, and these in a psychiatric rather than a department store setting. Preparation of the two O's for the present study consisted of their reviewing the structure of the partially standardized

interview and the rules for what constitutes scorable action and inaction. Experience had indicated that a major difficulty arises for the O when he tries to decide when an interviewee communication unit (action) has terminated and inaction has begun. In order to surmount this difficulty and thereby make the observations more objective, certain rules have been established to aid in the decision of what is scored as inaction. These rules have been published (23, pp. 362-364) and were reviewed by the two O's before the study began.

TABLE 4
OBSERVER RELIABILITY FOR TOTAL INTERVIEW†

	<i>Mean Raw Score</i>	<i>SD</i>	<i>rho</i>	<i>r</i>	<i>p</i>
1. S's Units					
Obs. X	50.70	18.92	.962	.985	.01
Obs. Y	50.70	18.79			
2. S's Action					
Obs. X	4153	847	.998	.998	.01
Obs. Y	4169	851			
3. S's Silence					
Obs. X	434	231	.910	.980	.01
Obs. Y	416	214			
4. S's Tempo					
Obs. X	4584	742	1.000	.999	.01
Obs. Y	4585	736			
5. S's Activity					
Obs. X	3722	990	.998	.996	.01
Obs. Y	3753	1000			
6. S's Adjustment					
Obs. X	—82	14.25	.710	.398	.01‡
Obs. Y	—5.41	22.78			
7. Int.'s Adjustment					
Obs. X	—85.41	76.03	.859	.944	.01
Obs. Y	—73.94	65.10			
8. S's Synchron.					
Obs. X	41.59	17.10	.928	.948	.01
Obs. Y	32.53	14.44			
9. Int.'s Units					
Obs. X	47.24	17.26	.993	.999	.01
Obs. Y	47.35	17.55			

† Note: "S" is Subject, "Int." is Interviewer, "Obs." is Observer.

‡ For this variable rho is significant at the .01 level, while r is not significant due to 3 deviant cases.

Following the joint review, standardized interviews were conducted by three experienced (employment) interviewers over a one-week period with seventeen randomly selected S's. The seventeen S's, who were being routinely interviewed and evaluated by the personnel department, consisted of applicants for jobs and employees being considered for promotion. The interviews were simultaneously but independently recorded by the two O's who sat in a small totally dark room and watched the interview through a one-way window. An intercommunication system was used to transmit the voices in all but the first three interviews, when mechanical failure forced the O's to use visual cues alone. Because of the darkness of the room, the use of earphones, and the observers' distance from the recording machines, no visual or auditory cues were available to either O to indicate the other's recording, assuring independence of the two sets of observations.

Table 4 contains the means and sigmas, as well as the Spearman rho and Pearson r reliability coefficients for nine major Interaction Chronograph variables. Although mean values (i.e., raw scores divided by number of units for each S) are usually used as scores for individual S's (as in all subsequent tables), it was felt that these might obscure differences between the results of the two O's. Therefore, individual raw scores were used in the computations. Mean values for the 17 S's (i.e., the means across the 17 individual raw scores) are presented for the total interview in Table 4. The reason for presenting values for both rho and r in Table 4 is made clear by the results with the variable S's Adjustment. For this variable, due to the influence of only 3 cases, r is reduced to an insignificant value of .398, while rho, a measure which is less sensitive to extreme deviations, yields a highly significant value of .710. Since we have been dealing with inter-

action measures the characteristics of which we are only now beginning to determine, and with relatively small N's, we felt it wise to compute both r and rho in all our analyses. Taken together, the results in Table 4 are striking evidence that even with an inexperienced observer, recordings of Interaction Chronograph patterns during standardized interviews are very reliable. With r as the measure of reliability, eight of the nine variables have coefficients above .94, while six are above .98. The results are equally striking with rho.

Since the standardized interview consists of five subperiods, it is of interest to ask how reliable are the observations for these periods in contrast to the interview as a whole. Table 5 presents the subperiod observer reliability for a sample of three of the major interaction variables: the number of S's Units of action, the duration of S's Actions, and the duration of S's Silences. The values of r within subperiods for these three variables range from .787 to .999, despite the fact that each is based on only a small time sample of the total interview. The one relatively low value of rho, .476, for S's Units in Period 2, is a statistical artifact due to a number of tied ranks, as can be seen by the high value (.889) of the Pearson r for this same variable. Of the 45 period-variable combinations (9 variables times 5 subperiods, of which

TABLE 5
OBSERVER RELIABILITY FOR
INDIVIDUAL PERIODS

Period	S's Units		S's Action		S's Silence	
	rho	r	rho	r	rho	r
1	1.000	.998	.958	.967	.910	.952
2	.476	.889	.988	.997	.855	.911
3	.972	.973	.946	.970	.781	.840
4	.710	.787	.994	.999	.763	.940
5	.984	.984	.892	.878	.901	.956

$p = .05$, $r = .482$, $\rho = .49$

$p = .01$, $r = .606$, $\rho = .64$

only 15 are shown in Table 5), 10 of the Pearson r observer-reliability values were .99; 20 were .95 and above; 28 were .90 and above; and 40 (89%) were 70 and above.¹ Similar values were found for the 15 rho coefficients shown in Table 5. Only one variable, S's Adjustment (in Periods 2 and 3) was found to be unreliable. These two instances of subperiod unreliability appear to be due in part to the very restricted number of observations relevant to S's Adjustment which occur in Periods 2 and 3. Since S can fail to adjust (interrupt or fail to respond) to the interviewer only when the interviewer himself has acted, the occurrence of approximately three and five interviewer's Units in Periods 2 and 3, respectively, meant that S's Adjustment in these periods depended upon very few (three and five) observations. Therefore, relatively small differences in observing one unit of adjustive behavior out of the three instances led to unreliability between O's for this variable during these two subperiods. Since these were the only two instances of unreliability, it can be concluded that observer

reliability is high for individual subperiods as well as for the interview as a whole.

Table 6 presents the reliability of the three variables which are scored only in the stress periods (Periods 2 and 4) of the standardized interview. These variables are S's Initiative, S's Dominance, and S's Quickness. These variables, like those in Table 4, are usually divided by the number of S's Units and hence express average frequency, or duration per unit. The individual raw scores were utilized in the present study, however, as stated earlier. It is clear from the values shown in Table 6 that, despite the fact that these three variables are derived from only a small sample of the total interview, the O's nevertheless attained considerable reliability (.01 level of confidence for S's Initiative and S's Quickness, and .05 level for S's Dominance). The finding of significant but lower observer reliability for the Period 4 S's Dominance variable would seem to support our earlier hypothesis (38, p. 427) that the "fast pace" of Period 4, with both the S and interviewer talking at the same time, presents the O with the most difficult recording situation. Reference to Table 5 of this study sheds further light on this possibility. In this table the Pearson r 's for S's Action and S's Silence

4. To save printing costs, these 45 period-variable correlations have been deposited with the American Documentation Institute. Order Document No. 5183, remitting \$1.25 for microfilm or \$1.25 for photocopies.

TABLE 6

OBSERVER RELIABILITY FOR INITIATIVE, DOMINANCE, AND QUICKNESS

	<i>Mean Raw Score</i>	<i>SD</i>	<i>rho</i>	<i>r</i>	<i>p</i>
1. S's Initiative (Period 2)					
Obs. X	7.47	2.81	.765	.877	.01
Obs. Y	8.53	3.39			
2. S's Dominance (Period 4)					
Obs. X	3.18	6.66	.532	.588	.05
Obs. Y	6.59	7.37			
3. S's Quickness (Period 2)					
Obs. X	—127.59	60.04	.941	.954	.01
Obs. Y	—128.82	55.34			

in Period 4 are extremely high (.999 and .940, respectively), while the r for S's Units is somewhat lower (.787) but still at the .01 level. Such results suggest that the O's differ by only several hundredths of a minute in recording how long an S speaks and is silent in Period 4, but that the extremely small differences in duration occasionally result in observer disagreements as to whether S stopped acting before or after the interviewer stopped acting and thus somewhat reduce observer reliability for S's Dominance. Likewise, the very small discrepancies in observer input for the duration measures may occasionally result in differences as to whether S stopped very briefly and then began a new unit or was continuously acting in one unit. As is shown in Table 5, however, such minor differences in observer-input for duration measures apparently have little effect on the reliability of some of the variables (S's Units, S's Silence, S's Action) even in Period 4, although they may result in more serious differences in the scores obtained for S's Dominance (Table 6). However, the fact that S's Dominance could yield an observer reliability coefficient at the .05 level of confidence despite the fast pace of Period 4 implies that, with further refinements in definition, and possibly more observer practice in recording Period 4 interactions, this variable may be as reliably observed and recorded as the others.

Considering the study as a whole, it is clear from the results presented that, with the possible exception of the S's Dominance variable, the observation and recording of interaction patterns during the partially standardized interview is a highly reliable undertaking. The unusually high coefficients of correlation for the total interview (Table 4) imply that the observer's task is largely a mechanical one once he has read, understood, and practiced the published rules as to what constitutes an action and an inaction (23,

pp. 362-364). Observer response-sets or biases appear to have little effect upon the interviewee interaction record finally obtained (30).

Scorer Reliability

Also to be examined was the reliability of the scorer who abstracts the objective scores from the numbers printed by the counters. The Interaction Chronograph yields cumulative scores on the several variables and thus scoring, which involves primarily simple arithmetic skills, is a reasonably objective procedure. Two scorers followed Chapple's manual of instructions (6) and scored independently the Interaction Chronograph records of ten standardized interviews selected at random. The results indicated perfect agreement between two scorers on 96 percent of the 600 individual final scores involved. The magnitudes of the errors involved in the remaining 4 percent were very minimal: they were of the order of 1 unit in whole number scores and .06 in those variables measured in hundredths of a minute. Thus it would appear that scorer reliability presents no problem in these observations.

Interviewer Reliability

By the nature of the standardized interview itself, interviewer and interviewee performance are mutually dependent and thus the question of the reliability of the interviewer's standardized performance (the independent variable) is confounded with that of the interviewee's performance (the dependent variable). Thus the patient, free to manifest his individuality, sets the pattern both for content and temporal characteristics, while the interviewer must follow him, imposing only the predefined constraints set forth in Tables 2 and 3. Indirect approaches to assessing the reliability of each participant are possible, however.

Reliability of the interviewer's behavior is indicated in the following several ways. First, the machine records for our

interviewers showed that, when they were first learning the standardized interview method, their behavior deviated widely and randomly from the rules. For example, durations of single utterances varied from 1 second to 18 seconds; instead of the prescribed 5 seconds; the number of failures to respond in Period 2, or interruptions in Period 4 varied widely on both sides of the stated 12; etc. Only when the machine records indicated closer adherence to the rules by each interviewer (approximately 40 practice interviews each) were the experiments to be described below undertaken. However, before these two-interviewer reliability studies are described, results on how well a single interviewer can follow the rules will be presented.

Because of the present form of the Interaction Chronograph, it has only recently become possible for us to examine the doctor-patient interactions unit by unit. This is primarily because certain variables are algebraic sums of positive and negative values. For example, Actions and Silences are not recorded as such but are summed under two counters; one for Activity (action minus si-

lence) and one for Tempo (action plus silence). Another reason is that certain variables in which we are interested (e.g., number of failures to respond and number of interruptions) are in no way recorded but can be determined only from a unit by unit analysis of single interview records. Without an analysis of these and other first-order variables also not recorded directly by the present form of the instrument, it is not possible to document the degree to which the interviewer in fact follows the rules of the standardized interview.

Although the necessary analytic procedures were tedious, we undertook the examination of the records of the first interviews of 10 patients randomly selected from a group interviewed at the Massachusetts General Hospital (in a study to be described below). The total number of interviewer units with these 10 patients was 529, and these constitute the data from which were obtained the results to be described below.

Table 7 shows the degree to which the interviewer is able to follow the rule that the duration of each of his actions should be 5 seconds. It is clear from Table 7

TABLE 7

DR.'S ACTION IN EACH OF THE FIVE PERIODS OF THE STANDARDIZED INTERVIEW:
MEANS, STANDARD DEVIATIONS AND RANGES

	<i>Period</i>					<i>F-test</i>	<i>p</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>		
<i>Dr.'s Action</i>							
Mean†	4.83	5.11	4.92	4.61	4.71	1.41	NS
SD†	1.51	1.67	1.41	1.37	1.54		
Range†	0.6-10.8	1.8-9.0	1.8-9.0	1.2-9.0	1.2-12.6		
N (Dr.'s Units)‡	155	40	99	133	102		

† Values shown are in seconds.

‡ The number of units upon which the means for these 10 patients are based vary from period to period due to (1) differences in the *lengths* of each of the periods, (2) individual differences among patients, and, most importantly, (3) the fact that in most circumstances the Dr. is not to talk in Period 2 except when the patient has not taken the initiative by speaking again within 15 seconds following his own last utterance.

N = 529

that his actions averaged almost exactly 5 seconds in each of the 5 periods. The ranges indicate that there was an occasional departure from the rule. However, analysis revealed that 64 percent of the doctor's individual actions were between 4 and 6 seconds, and that 90 percent of them were between 3 and 7 seconds. Since another rule for the interviewer is that he must stop within about 2 seconds whenever he is interrupted by the patient, a small number of interviewer actions is expected to be under 5 seconds.

Table 8 shows the degree to which the interviewer is able to follow the rule that the duration of each of his silences in Periods 1, 3, and 5 should be less than 1 second. The average duration is approximately 0.64 second, as shown in Table 8. Further analysis showed that 97 percent of the interviewer's individual silences were under 1 second.

Table 9 shows the degree to which the interviewer is able to follow the rule that for those instances in Period 2 when the patient remains silent following his own last utterance (i.e., fails to initiate by speaking again), the interviewer is to wait 15 seconds before speaking himself. The mean (16.84 seconds) and range (11.4 to 20.4 seconds) indicate a fairly high degree of success. Eighty percent of the doctor's required 15 second silences fell between 15 and 19 seconds.

Table 10 shows the degree to which the interviewer is able to follow the rules

that he fail to respond to the patient's last utterance 12 times in Period 2, and that he interrupt the patient 12 times in Period 4. In both instances the mean of 12 and range of 11 to 13 indicate that the interviewer follows these two rules without difficulty.

These tables make clear that the interviewer is able both to learn and to follow the rules of the standardized interview to a reasonably high degree. With the demonstration that learning to follow the rules of the standardized interview is possible, we can next turn to another aspect of the interviewer reliability problem. Namely, how reliably can two interviewers carry out the standardized interview with a given sample of subjects?

When two interviewers had independently learned the standardized interview to a satisfactory criterion, and had then interviewed 20 outpatients in our first interviewee reliability study, observations on this aspect of the interviewer reliability problem were made. This study was designed to investigate primarily the reliability of interviewee interaction patterns. One aspect of this study pertains to the question of interviewer reliability and will be presented here. Each patient was interviewed by each interviewer in interviews separated by an interval of a few minutes, in ABBA order and independently.

That interviewer reliability for another interview dimension is high is shown in

TABLE 8

DR.'S SILENCE IN THREE FREE PERIODS OF THE STANDARDIZED INTERVIEW:
MEANS, STANDARD DEVIATIONS AND RANGES

	<i>Period</i>			<i>F-test</i>	<i>p</i>
	<i>1</i>	<i>3</i>	<i>5</i>		
1. Mean Dr.'s Silence†	0.64	0.66	0.61	3.33	NS
SD†	0.21	0.17	0.07		
Range†	0.6 to 1.8	0.6 to 1.8	0.6 to 1.2		
N (Dr.'s Units)	101	67	71		

† Values shown are in seconds.

Table 11, which deals with the lengths of the pair of interviews for any one subject. From Table 2 it can be seen that twenty minutes of the interview are fixed and that, depending upon the subject's behavior in Periods 2 and 4, anywhere from several to thirty more minutes may be utilized by a subject to complete the standardized interview. Table 11 contains an analysis of the actual mean length of the twenty pairs of interviews. It can be seen that the standardized interview takes, on the average, about 33 minutes to conduct, and that both doctors had essentially equal interview times. Furthermore, the order of the interview, comparing all 20 *first* interviews or all 20 *second* interviews—independent of which doctor was first or second—yielded no differences in mean length of interview when subjected to statistical analysis (analysis of covariance). It is of interest to remember that relatively little restriction was placed on the lengths of two of the periods (i.e., patients could vary in their behavior in Periods 2 and 4 such that the lengths of each of these periods could range from, say, one minute to fifteen minutes). Table 11 indicates that, even with this freedom to vary, the patients behaved similarly in regard to this variable with the two interviewers. Although not shown in Table 11, the mean lengths of the two variable subperiods were similar. They were respectively 8.41 and 8.44 minutes in Period 2 ($r = .631$), and 3.68 and 2.26 minutes in Period 4 ($r = .555$). These slight differ-

TABLE 9

DR.'S SILENCE IN THE SILENCE STRESS PERIOD (PERIOD 2) OF THE STANDARDIZED INTERVIEW: MEANS, STANDARD DEVIATIONS, AND RANGES

<i>Period 2</i>	
1. Mean Dr.'s	
Silence (in seconds)	16.84
SD (in seconds)	2.25
Range (in seconds)	11.4 to 20.4

TABLE 10

DR.'S FREQUENCY OF FAILURES TO RESPOND (PERIOD 2) AND FREQUENCY OF INTERRUPTIONS (PERIOD 4): MEANS, STANDARD DEVIATIONS AND RANGES

	<i>Failures to Respond</i>	<i>Inter- ruptions</i>
Mean number	12.20	12.00
SD	0.60	0.15
Range	11 to 13	11 to 13

ences in mean lengths were not significant by F-test, while the values of r were significant at the .01 level of confidence. These indirect observations, although confounded by interviewee characteristics, point to the reliability of one further aspect of the interviewer's behavior during the standardized interview.

Similarly, the high agreement in the mean number of *Dr.'s Units*, when the same 20 patients were independently interviewed by the two doctors, points to interviewer reliability on still another dimension. The data are shown in Table 12.

It is clear from both Tables 7, 8, 9, and 10, and additionally from Tables 11 and 12, that the ability to learn and follow the rules is not limited to one interviewer. However, it is also clear that trained interviewers, while quite reliable, are not perfect. We know of several possible sources of the slight variability documented in the above tables; fluctuations (error) in the interviewer's performance, on the one hand, and on the other hand, and probably more important, his difficulty in deciding quickly, from a combination of verbal and non-verbal cues, when the patient has terminated an action, when the patient is about to initiate a new action, whether in the rapid single exchanges in Period 4 the patient has, in fact, stopped acting or not, etc. These difficulties notwithstanding, Tables 7 to 12 show that the inter-

TABLE 11
ANALYSIS OF MEAN LENGTH OF TOTAL INTERVIEW
FOR 20 PATIENTS INTERVIEWED TWICE

<i>Analysis</i>	<i>Mean</i>	<i>Range</i>
Length of all 40 interviews	32.8 minutes	25.7 to 50.3
a. Length of all first interviews (N = 20)	32.9 minutes	26.9 to 41.4
b. Length of all second interviews (N = 20)	32.8 minutes	25.7 to 50.3
c. Length of all first doctor's interviews (N = 20)	33.5 minutes	26.9 to 50.3
d. Length of all second doctor's interviews (N = 20)	32.2 minutes	25.7 to 41.2

viewer can, with a little practice, become an "instrument" of research of considerable reliability.

Interviewee Reliability: First Study

After two interviewers had practiced the interview technique described in Tables 2 and 3 in some 40 interviews each, and the objective interaction scores indicated that they were approaching the prescribed patterns with a high degree of accuracy and without feeling unduly inhibited in their responses, the first of several studies of interviewee reliability was initiated (24, 38, 39).

The design called for twenty patients, each to be interviewed *separately* by the two interviewers, one of whom was a young internist experienced in interviewing, and the other an older psychiatrist. The Ss were out-patients sent to the psychiatric clinic of Washington University School of Medicine for screening and assignment to individual psychotherapists. They were new patients, white, 11 men and 9 women, ranging in age from approximately 18 to 55. The pre-

sending problems were typical of the population of this outpatient clinic, and consisted of cases of anxiety reaction, hysteria, depression, schizophrenia, obsessive-compulsive neurosis, duodenal ulcer, and possible chronic brain syndrome.

The design was randomized to control order effects and each interviewer thus interviewed every other patient first; an AB order with the first patient and a BA order with the second, etc. In this way each interviewer came first in the sequence for ten patients and second for the remaining ten. Statistical evaluation of "interviewer-order" effects on patients' behavior was therefore possible. This proved to be an unimportant variable (24).

Ash trays were removed from the interviewer's desk so that patient smoking would be discouraged and thus not complicate the observer's task. No mention was made of the experiment by the interviewer and the experiment began when he opened the interview with such a five-second statement as "My name is Dr.—"

TABLE 12
DR.'S UNITS FOR TWO INTERVIEWERS CONDUCTING INDEPENDENT
INTERVIEWS WITH 20 PATIENTS A FEW MINUTES APART

	<i>Original Study</i>		<i>r</i>	<i>rho</i>	<i>p</i>
	<i>Dr. 1</i>	<i>Dr. 2</i>			
1. Mean Dr.'s Units	63	60	.772	.782	.01
SD	23	18			
Range	21 to 109	27 to 93			

Can you tell me how you happened to come to the clinic at this time?" The interviewer pressed a concealed button for a light signal to the observer. The observer was seated in the next room on the other side of a large one-way window (5 ft. x 2 ft.) and recorded the interaction by pressing designated buttons for the interviewer and the patient. A high-fidelity microphone was hung from the ceiling and connected to an amplifier in the observer's room. The side effects of the recorded voices in the observer's room were controlled by piping the voices to the observer through earphones. Throughout the experiment, sound recording and visual observation of the interactions were excellent. Verbatim recording of the interview exchanges was made by means of recently devised highly sensitive microphones, connected directly to an Audograph machine in the observer's room. The Audograph discs were transcribed for subsequent studies (content analysis, etc.).

The second interviewer was always in another building when the first interviewer was conducting his interview, thereby insuring independence: both began with no knowledge about the patient. After finishing his interview, the first interviewer would say, "Mr(s).—, I'd like another doctor in our clinic to talk with you now, and if you will wait here a minute or two I will go to get him." Upon his arrival, the second interviewer signalled the observer, made a five-second opening introductory remark to the patient, and the second half of the experiment began.

Subjects were used in the order in which they were sent to this particular clinic. However, due to power failure in the hospital on one occasion, and to total incoherence on the part of one patient, two of the consecutive planned experiments were lost.

Table 13 presents the means, standard deviations, and ranges for eight variables

which are recorded for the total interview, while Table 14 presents similar data for the two variables recorded only in Periods 2 and 4. The data for two of the remaining variables are presented in Tables 11 and 12. The results show a striking reliability (stability) in these interviewee interaction variables from first interview to second interview, for each of the twenty patients. The stability in patients' interaction patterns (and corollary reliability of measurement by the Interaction Chronograph) is demonstrated in Table 13 in several ways: (a) *relative* reliability of the interaction behaviors manifested by each of the 20 patients with the two doctors, and (b) *absolute* reliability of these same interaction patterns.

Relative reliability is inferred from the values of the Pearson product-moment coefficients of correlation shown in the second column from the last. These coefficients range from .726 to .930 and all are significant at the .01 level of probability. They indicate that, for the interview as a whole (summing all five periods) each of the 8 Interaction Chronograph variables has a marked stability, i.e., that two interviewers will elicit interviewee characteristics from each of the 20 patients in amounts which place each patient in the same relative position.

However, since one can get high values of Pearson r for any one variable when 2 interviewers elicit *different amounts* of this variable from the same patient, providing the same *relative* values of these amounts are maintained from patient to patient,⁵ these high values for r , in themselves, are not sufficient to demonstrate

5. One could get a perfect correlation (1.00) on Pt.'s Units, for example, if the second interviewer *always* got double or some other multiple of the number of units from the same patients; i.e., if from 3 Pts. one doctor got 30, 40, and 50 units, and the second obtained 60, 80, and 100, the value of r would be 1.00.

TABLE 13

MEANS, STANDARD DEVIATIONS, RANGES AND COEFFICIENTS
OF CORRELATION ACROSS TOTAL INTERVIEW

		<i>Original Study</i>		<i>r</i>	<i>p</i>
		<i>Dr. 1</i>	<i>Dr. 2</i>		
1	Mean Pt.'s Units	72.20	69.85	.747‡	.01
	SD	25.55	20.90		
	Range	25 to 127	29 to 112		
2.	Mean Pt.'s Action†	48.20	43.90	.912	.01
	SD	39.64	29.84		
	Range	13 to 154	9 to 136		
3.	Mean Pt.'s Silence†	9.10	8.20	.764	.01
	SD	3.85	2.46		
	Range	4 to 19	4 to 13		
4.	Mean Pt.'s Tempo†	57.20	52.85	.908	.01
	SD	36.87	27.26		
	Range	24 to 159	26 to 142		
5.	Mean Pt.'s Activity†	39.55	36.20	.930	.01
	SD	40.49	30.01		
	Range	—2 to 150	2 to 131		
6.	Mean Pt.'s Adjustment†	—1.53	—1.28	.802	.01
	SD	1.33	1.17		
	Range	0 to —5	0 to —4		
7.	Mean Dr.'s Adjustment†	—1.83	—1.53	.737	.01
	SD	1.17	.80		
	Range	—3.57 to +1.00	—3.03 to +.14		
8.	Mean Pt.'s Synchronization	.84	.84	.726	.01
	SD	.07	.06		
	Range	.63 to .99	.76 to .98		

† Values for these variables are recorded in hundredths of a minute. To convert to seconds multiply the given value by 0.6.

‡ Corresponding values for rho for these eight variables are as follows: (1) .807, (2) .847, (3) .854, (4) .805, (5) .874, (6) .716, (7) .668, and (8) .691.

TABLE 14

MEANS, STANDARD DEVIATIONS, RANGES, AND COEFFICIENTS OF CORRELATION FOR TWO
INTERVIEWERS FOR INITIATIVE (PERIOD 2) AND DOMINANCE (PERIOD 4)

		<i>Original Study</i>		<i>r</i>	<i>rho</i>
		<i>Dr. 1</i>	<i>Dr. 2</i>		
1.	Mean Pt.'s Initiative (Period 2)	.75	.77	.805**	.718**
	SD	.18	.19		
	Range	.17 to .94	.23 to 1.00		
2.	Mean Pt.'s Dominance (Period 4)	— .32	— .42	.470*	.102
	SD	.31	.30		
	Range	— .71 to +.83	— .92 to +.29		

* Significant at the .05 level of probability.

** Significant at the .01 level of probability.

reliability in the sense of stability (or invariance) in patient characteristics.

To demonstrate the latter, one needs to examine the values of the means, SD's and ranges for any one variable across the two interviewers. Examination of Table 13 (and Table 12) indicates equally striking *absolute* reliability of these 9 variables for the 20 patients with the two interviewers. Thus, for example, despite very marked individual differences in the interaction rate of the 20 patients (Pt.'s Units), as seen by the fact that patients varied from 25 to 127 units with Doctor 1, and 29 to 112 with Doctor 2, the average number of units for these 20 patients was 72 with Doctor 1 and almost 70 with Doctor 2. That is, despite the fact that, during an interview of 33 minutes' average length, one patient interacted 25 times (less than once a minute) and another patient 127 times (4 times per minute) with Doctor 1, and similarly with Doctor 2, any *single* individual seemed to maintain his *own* interaction rate independent of which doctor was interviewing him. Thus one individual interacted 25 times with the first doctor and 29 times with the second, while another subject interacted 112 and 129 times, respectively, with the two interviewers. This marked similarity in values of the means, as well as the equally striking similarity in standard deviations and ranges for all the variables, would tend to indicate, when added to the high values of the Pearson *r*'s, that the interaction variables reflect stable and invariant personality characteristics under the standardized conditions in the subjects studied.

While the results shown in Table 13 (and Table 12) are the reliability figures for the total interview, the reliability of the same variables in each of the five subperiods was likewise high. These data have been published elsewhere (24, 38, 39), and will not be presented here be-

cause to date very little use has been made of the subperiod scores. However, there are two variables, Pt.'s Initiative and Pt.'s Dominance, which are scored only in the two stress periods (Periods 2 and 4). Table 14 contains the test-retest reliability data across the two interviewers for these two variables in the same sample of 20 outpatients. Two things are apparent from the findings presented in Table 14: (1) the Pearson Coefficients of Correlation are statistically significant, even though the average Pt.'s Initiative score is derived from Period 2, which has an average length of only about 8 minutes of the total interview average of 32 minutes, and the average Pt.'s Dominance score from Period 4, which takes only about 3-4 minutes of the total 32; (2) for this group of 20 subjects, the reliability of the Pt.'s Dominance variable (.470) is lower than that of the Pt.'s Initiative variable (.805). One possible explanation of the lower reliability of the Pt.'s Dominance variable, as an interviewee characteristic, is the lower reliability of the observer's input for this variable, as shown in Table 6. However, the results presented in Tables 16 and 18 indicating as they do a reversal in the relative stability of these interaction variables, will make evident the fact that it probably is not observer error alone which is responsible for the lower reliability value for Dominance (shown in Table 14), but that more likely this reduced reliability is a reflection of the fact that Dominance is assessed for only 3-4 minutes of the total 32 minute (average) interview. Thus, the reduced reliability may be merely a reflection of the Spearman-Brown phenomenon in reverse (that a sample of a whole will be less reliable than is the longer whole, itself). This phenomenon would hold also for Pt.'s Initiative (based on 8 minutes) relative to the other interviewee variables all of which are assessed on the basis of the total interview (Table 13).

Interviewee Reliability: Replication Study

Since the reliabilities obtained for the Interaction Chronographs variables were unusually high for complex variables such as one finds in the interview, it was decided to duplicate the study exactly with a second sample of 20 patients referred to the same clinic. All conditions of the replication study were the same as before. The results are shown in Tables 15 and 16. As will be seen, they confirm the findings of the first study.

The results so far described (in Tables 4 through 16) thus define (1) the reliability of the interviewer who serves as the independent variable by following the rules of the partially standardized interview; (2) the reliability of the interviewee interaction patterns, (the dependent variables); (3) the reliability of the scorer who scores the final Interaction Chronograph record; and, finally, (4) the reliability of the observer's input.

POSSIBLE CONSTRAINING INFLUENCE OF THE STANDARDIZATION OF THE INTERVIEW

The question can be raised here whether the obtained high stability in interaction patterns from interviewer to interviewer for any one interviewee was due to the latter's repeating the "same story" to the second interviewer, or some similar artifact. Verbatim records of the interviews show that there were often major differences in content between the first and second interviews. Military experiences might be a prominent feature of the first interview, yet hardly alluded to in the second: resented desertion by a spouse might dominate one or more subsections of the first interview, and not be mentioned in the second, etc. Hence the "same story" possibility is not a likely explanation of the results.

The records show also that psychotherapeutic changes could occur in the first interview and be reported in the second, prefaced by a clearly identifying

remark such as "Until I talked to the other doctor today, I didn't realize that . . .," etc. Some patients were obviously less emotional in the second interview than in the first. Such differences between the two consecutive interviews in content, attitude, insight, emotionality, etc., are material for examination by a variety of methods. However, in spite of these clinically obvious differences in the two interviews of one patient, there was still a marked stability in interaction patterns for any given patient.

Some post-interview comments made either to one of the interviewers or to a doctor in another clinic, such as "I liked you more than the other doctor—I didn't like him at all," "I was upset by those doctors—I revealed things and said things I hadn't thought of for years," etc., emphasize that: (a) the standardized interview is compatible with certain patient reactions common to any initial psychiatric interview; (b) change in the interactional pattern of the interviewer from period to period is not mentioned by the patient who presumably is not fully aware of it; and that therefore (c) there exists the possibility of utilizing the standardized interview as a systematic observational device that can be interpolated a number of times at different points in a prolonged psychotherapeutic or other interactional relationship, without "accommodation" or practice effects.

STABILITY IN THE INTERVIEW BEHAVIOR OF SUBJECTS RE-INTERVIEWED AFTER VARYING INTERVALS

The above observations deal with the stability of the behavior of subjects independently interviewed twice in the same afternoon (a few minutes apart) by two practiced interviewers. Having demonstrated that, on a given day two interviewers could each elicit the same interaction behavior from any given patient, we examined the stability of the indi-

TABLE 15

MEANS, STANDARD DEVIATIONS, RANGES, AND COEFFICIENTS
OF CORRELATION ACROSS TOTAL INTERVIEW

	<i>Replication Study</i>		<i>r</i>	<i>p</i>
	<i>Dr. 1</i>	<i>Dr. 2</i>		
1. Mean Pt.'s Units	68.30	76.65	.926‡	.01
SD	25.95	26.69		
Range	39 to 132	41 to 133		
2. Mean Pt.'s Action†	44.15	39.35	.899	.01
SD	23.08	64.26		
Range	12 to 93	12 to 93		
3. Mean Pt.'s Silence†	9.10	9.00	.770	.01
SD	2.92	2.25		
Range	5 to 18	6 to 15		
4. Mean Pt.'s Tempo†	53.65	48.55	.905	.01
SD	22.33	19.61		
Range	19 to 100	19 to 100		
5. Mean Pt.'s Activity†	35.00	30.35	.891	.01
SD	23.35	20.28		
Range	4 to 87	3 to 85		
6. Mean Pt.'s Adjustment†	-1.08	-.66	.853	.01
SD	1.43	1.29		
Range	+.50 to -5.64	+.88 to -4.95		
7. Mean Dr.'s Adjustment†	-2.24	-1.93	.749	.01
SD	1.21	.69		
Range	-.77 to -4.82	-.88 to -2.97		
8. Mean Pt.'s Synchronization	.84	.85	.741	.01
SD	.08	.05		
Range	.69 to .97	.77 to .98		
9. Mean Dr.'s Units	58.75	63.45	.909	.01
SD	26.16	24.94		
Range	28 to 119	32 to 115		

† Values for these variables are recorded in hundredths of a minute. To convert to seconds multiply the given value by 0.6.

‡ Corresponding values for rho for these nine variables are as follows: (1) .917, (2) .945, (3) .859, (4) .941, (5) .910, (6) .821, (7) .780, (8) .717, and (9) .880.

TABLE 16

MEANS, STANDARD DEVIATIONS, RANGES, AND COEFFICIENTS OF CORRELATION FOR
TWO INTERVIEWERS FOR INITIATIVE (PERIOD 2) AND DOMINANCE (PERIOD 4)

	<i>Replication Study</i>		<i>r</i>	<i>rho</i>
	<i>Dr. 1</i>	<i>Dr. 2</i>		
1. Mean Pt.'s Initiative (Period 2)	.72	.73	.552*	.556*
SD	.18	.14		
Range	.25 to .92	.42 to .92		
2. Mean Pt.'s Dominance (Period 4)	-.36	-.43	.697**	.713**
SD	.31	.26		
Range	-1.00 to +.13	-.92 to +.00		

* Significant at the .05 level of probability.

** Significant at the .01 level of probability.

vidual subject's interview behavior over time. In order to control for possible interviewer interaction effects (interaction in the statistical sense of confounding variables) only one interviewer was used for both test and retest. It was recognized, however, that the use of one interviewer for both interviews, while avoiding the difficulty of interviewer confounding, nevertheless might introduce other difficulties involving habituation, "transference," or other patient "set" responses—and thus produce lower reliability figures on retest.

For each of the studies to be presented, the method of procedure was as described earlier. Each subject was interviewed twice by one interviewer, the same experienced psychiatrist used in the earlier studies. The retest intervals used were three: (a) 7 days, (b) 5 weeks, and (c) 8 months.

Tables 17 and 18 show the stability of interview interaction patterns over seven days in a third sample of 20 out-patients from the Washington University School of Medicine psychiatric outpatient clinic.

Tables 19 and 20 show the stability of interview interaction patterns (over 5 weeks) of a fourth sample of patients.⁶ These were 19 white, male, chronic schizophrenic patients ranging in age from 27 to 42 years (median of 30 and mean of 33), who were interviewed at the Research Facility of the Rockland State Hospital.⁷ Diagnosis was made and agreed upon by two staff psychiatrists in

every instance. The patients had been hospitalized (uninterruptedly) with this diagnosis for a minimum of three years (range from 3 to 18 years, with a mean of 8.3, and median of 7.5 years). Mean and median years of education were 9.7 and 9.2 respectively, with a range from 4 to 16 years.

It is clear from Tables 17, 18, 19, and 20 that considerable stability of the interviewee interaction variables is still evident after a 7-day and a 5-week interval. The method thus can be used as a systematic observational device which can be interpolated a number of times at different points for the study of behavioral changes over time (e.g., for studying the effects of psychotherapy, drugs, psychosurgery, etc.).

Reliabilities which, after one week (Tables 17 and 18) are lower than after five weeks (Tables 19 and 20), raise interesting questions. It may be that a hospitalized, chronic schizophrenic population is more stable in its interactional characteristics than is an outpatient, non-psychotic group of patients. It is possible, also, that the characteristics of the interviewee population are unimportant in themselves and that the relevant variable is the time interval between the interviews. That is, "memory," "habituation," "transference," etc. phenomena all may be important here. It may turn out that the preservation of the "newness" of the stimulus situation (medical atmosphere, the particular interviewer as a person, the patient's drive state at the time of interview, etc.) is highly pertinent to the degree of stability of the interviewee's interaction pattern. A third possible explanation is that the lower stability after 7 days may be a statistical artifact: the standard deviations for nearly all the interaction variables are much smaller for the 7 day group (Table 17) than for the same day (Tables 12, 13, 15) or the 5 week group (Table 19). Choice among

6. These observations were made by use of a portable tape-recording device, the information in which could subsequently be fed into the standard form of the Interaction Chronograph. The portable device was developed by the E. D. Chapple Company, Inc., of Noroton, Connecticut.

7. We should like to thank Alfred M. Stanley, M.D., Nathan S. Kline, M.D., and other members of the medical, nursing, and research staff for the considerable time and effort they contributed on our behalf in the execution of this study.

TABLE 17

MEANS, STANDARD DEVIATIONS, RANGES AND COEFFICIENTS
OF CORRELATION ACROSS TOTAL INTERVIEW

	<i>One Week Apart</i>		<i>rho</i>	<i>r</i>
	<i>First Interview</i>	<i>Second Interview</i>		
1. Mean Pt.'s Units	72.30	78.30	.765**	.890**
SD	15.75	19.54		
Range	43 to 118	48 to 131		
2. Mean Pt.'s Action	34.55	34.05	.597 [†] *	.655* [†]
SD	15.77	15.32		
Range	5 to 86	9 to 73		
3. Mean Pt.'s Silence	9.75	9.30	.532 [†] *	.646* [†]
SD	3.73	2.43		
Range	6 to 24	6 to 15		
4. Mean Pt.'s Tempo	46.00	44.40	.735 [†] *	.764* [†]
SD	14.19	14.59		
Range	27 to 96	22 to 84		
5. Mean Pt.'s Activity	28.00	25.75	.738**	.768**
SD	14.19	13.60		
Range	8 to 76	3 to 62		
6. Mean Pt.'s Adjustment	— .89	— 1.15	.776**	.757**
SD	.77	1.39		
Range	+ .58 to — 2.50	+ .89 to — 5.24		
7. Mean Dr.'s Adjustment	— 2.09	— 2.17	.558**	.554*
SD	.77	.87		
Range	— .98 to — 3.34	— .95 to — 3.69		
8. Mean Pt.'s Synchronization	.86	.88	.541 [†] *	.526**
SD	.05	.07		
Range	.76 to .93	.73 to 1.01		
9. Mean Dr.'s Units	59.65	66.25	.528 [†] *	.823* [†]
SD	13.88	18.49		
Range	37 to 96	37 to 118		

* Significant at the .05 level of probability.

** Significant at the .01 level of probability.

TABLE 18

MEANS, STANDARD DEVIATIONS, RANGES, AND COEFFICIENTS OF CORRELATION FOR
TWO INTERVIEWERS FOR INITIATIVE (PERIOD 2) AND DOMINANCE (PERIOD 4)

	<i>One Week Apart</i>		<i>r</i>	<i>rho</i>
	<i>First Interview</i>	<i>Second Interview</i>		
1. Mean Pt.'s Initiative (Period 2)	.70	.67	.335	.416*
SD	.19	.16		
Range	.29 to .92	.38 to .92		
2. Mean Pt.'s Dominance (Period 4)	— .46	— .36	.367	.404*
SD	.26	.35		
Range	0 to — .92	+ .75 to — .80		

* Significant at the .06 level of probability.

TABLE 19
MEANS, STANDARD DEVIATIONS, RANGES AND COEFFICIENTS
OF CORRELATION ACROSS TOTAL INTERVIEW

	<i>Five Weeks Apart</i>				
	<i>First</i>	<i>Second</i>			
	<i>Interview</i>	<i>Interview</i>	<i>r</i>	<i>rho</i>	<i>p</i>
1. Mean Pt.'s Units	80.95	73.05	.910	.907	.01
SD	34.93	32.65			
Range	24 to 139	25 to 152			
2. Mean Pt.'s Action	51.63	57.84	.862	.886	.01
SD	67.65	70.61			
Range	6 to 242	6 to 234			
3. Mean Pt.'s Silence	10.53	10.92	.779	.718	.01
SD	4.93	5.38			
Range	5 to 18	4.5 to 24.5			
4. Mean Pt.'s Tempo	62.68	68.79	.857	.919	.01
SD	65.24	68.25			
Range	18 to 248	18 to 239			
5. Mean Pt.'s Activity	40.95	46.89	.865	.874	.01
SD	70.39	73.29			
Range	-12 to +237	-15 to +230			
6. Mean Pt.'s Adjustment	-3.36	-4.02	.607	.530*	.01
SD	3.44	4.34			
Range	-9.78 to +.41	-15.98 to -.26			
7. Mean Dr.'s Adjustment	-1.46	-1.51	.737	.758	.01
SD	1.27	1.17			
Range	-2.84 to +1.25	-3.04 to +1.33			
8. Mean Pt.'s Synchron.	.97	.95	.596	.549*	.01
SD	.10	.10			
Range	.83 to 1.21	.80 to 1.21			
9. Mean Dr.'s Units	73.53	66.47	.890	.881	.01
SD	35.64	36.34			
Range	22 to 131	20 to 151			

* Significant at the .05 level of probability.

TABLE 20
MEANS, STANDARD DEVIATIONS, RANGES, AND COEFFICIENTS OF CORRELATION FOR
TWO INTERVIEWERS FOR INITIATIVE (PERIOD 2) AND DOMINANCE (PERIOD 4)

	<i>Five Weeks Apart</i>				
	<i>First</i>	<i>Second</i>			
	<i>Interview</i>	<i>Interview</i>	<i>r</i>	<i>rho</i>	<i>p</i>
1. Mean Pt.'s Initiative (Period 2)	.59	.64	.828	.812	.01
SD	.29	.29			
Range	0.00 to .92	.08 to .93			
2. Mean Pt.'s Dominance (Period 4)	-.67	-.71	.232	.143	NS
SD	.24	.20			
Range	-1.00 to -.20	-1.00 to -.11			

these alternatives can be made only after further investigation.

A fifth sample of patients was used to study stability of the interaction patterns over a still longer period of time. These were 20 white patients from the Massachusetts General Hospital, of whom 15 were outpatients and 5 inpatients; 8 males and 12 females. They had a mean and median age of 42.5 years (range from 15 to 80); median education of 10 years (range from 6 to 15); and median Wechsler-Bellevue I.Q. 101 (range from 73 to 124). Clinical diagnoses as given by one experienced psychiatrist were as follows: 8 depression, 6 anxiety neurosis, 2 hysteria, and one each of schizoid personality, psychopathic personality, unresolved grief reaction and possible psychosis (schizophrenia). The same psychiatrist used previously was the interviewer; the same observer was used; and in other ways the procedure was identical with that of the 7-day and 5-week studies. The standard Interaction Chronograph was used to record the interviews.

Tables 21 and 22 show the stability of interviewee interaction patterns in this sample over a median test-retest interval of 8 months (mean of 7.5 months and range from 3 to 12 months). It would appear from inspection of Tables 21 and 22 that the stability values are, for the most part, lower after 8 months than after shorter intervals.

In view of the fact that the time interval between test and retest interviews was so long (median of 8 months), it was not possible in this study to control possibly relevant variables as well as had been done with the previous time intervals used. Thus we learned, when our experiment was over, that the 20 patients had had varying numbers of psychotherapy hours between the test and retest standardized interviews. Furthermore, the various individual psychotherapists (who were not a part of our experiment) were of several different psychotherapeutic

schools, varied from inexperienced to very experienced, and had been assigned to these patients in terms of their medical and psychiatric needs, i.e., with no reference at all to our study. In this sense, the results to be presented below can be considered as arising from an "experiment of nature"

A SUGGESTED RELATION BETWEEN PSYCHOTHERAPY HOURS AND CHANGE IN INTERACTION BEHAVIOR

As we considered the results presented in Tables 21 and 22, and the facts just mentioned, it occurred to us that the lower reliability figures after 8 months might reflect a number of uncontrolled variables such as (1) unreliability of interviewee interaction patterns after this long an interval, (2) the complex variable "number of psychotherapy hours" interpolated between test and retest; and (3) other variables not clearly evident to us.

Careful inspection of the interaction records indicated that approximately half of the twenty patients had almost identical interviewee interaction patterns on retest, while the others showed considerable change. Before this fact could be demonstrated quantitatively, an index of change from first to second interview had to be determined. We used the following simple rule of thumb. For each patient, we computed the difference between his Units score in the first and second interview. These differences were rank-ordered from 1 to 20. A similar rank from 1 to 20 was obtained for each patient on the magnitude of the difference between his average Action in the two interviews. The two ranks (one for Units, one for Action) were then added for each patient. This sum then constituted the data from which was determined whether an individual was in the High Change Index group or Low Change Index group. The 10 individuals whose summed Units and Action ranks were the

TABLE 21

MEANS, STANDARD DEVIATIONS, RANGES, AND COEFFICIENTS
OF CORRELATION ACROSS TOTAL INTERVIEW

	<i>8 Months Apart</i>		<i>r</i>	<i>rho</i>
	<i>First Interview</i>	<i>Second Interview</i>		
1. Mean Pt.'s Units	90.15	87.75	.618 ⁺	.653 ⁺⁺
SD	27.90	27.69		
Range	44 to 137	45 to 147		
2. Mean Pt.'s Action	30.05	29.00	.475 ⁺	.665 ⁺⁺
SD	20.05	14.84		
Range	11 to 84	11 to 69		
3. Mean Pt.'s Silence	8.84	8.54	.598 ⁺⁺	.551 [*]
SD	1.64	1.65		
Range	6.2 to 13.7	6.0 to 11.1		
4. Mean Pt.'s Tempo	38.85	37.55	.458 ⁺	.625 ⁺⁺
SD	20.37	14.43		
Range	19 to 93	18 to 76		
5. Mean Pt.'s Activity	21.10	20.45	.485 ⁺	.596 ⁺⁺
SD	19.75	15.29		
Range	3 to 74	0 to 62		
6. Mean Pt.'s Adjustment	— .45	— .67	.560 ⁺⁺	.485 [*]
SD	.78	.77		
Range	—2.41 to +.60	—2.12 to +.68		
7. Mean Dr.'s Adjustment	—2.03	—1.99	.423 ⁺	.488 [*]
SD	.91	.68		
Range	—4.74 to —1.01	—2.97 to —.65		
8. Mean Pt.'s Synchronization	.88	.88	.740 ^{**}	.718 ^{**}
SD	.05	.07		
Range	.74 to .95	.73 to .96		
9. Mean Dr.'s Units	72.60	68.80	.665 ^{**}	.642 ^{**}
SD	26.24	25.93		
Range	29 to 119	28 to 116		

* Significant at the .05 level of probability.

++ Significant at the .01 level of probability.

TABLE 22

MEANS, STANDARD DEVIATIONS, RANGES AND COEFFICIENTS OF CORRELATION FOR
TWO INTERVIEWERS FOR INITIATIVE (PERIOD 2) AND DOMINANCE (PERIOD 4)

	<i>8 Months Apart</i>		<i>r</i>	<i>rho</i>
	<i>First Interview</i>	<i>Second Interview</i>		
1. Mean Pt.'s Initiative (Period 2)	.73	.70	.586 ^{**}	.656 ^{**}
SD	.17	.17		
Range	.31 to .93	.36 to .92		
2. Mean Pt.'s Dominance (Period 4)	— .55	— .66	.278	.151
SD	.20	.20		
Range	— .86 to — .11	— .92 to — .33		

** Significant at the .01 level of probability.

highest (11 to 20), were placed in one group, while the individuals whose summed Unit and Action ranks were the lowest (1 to 10), were placed in the other group. The test-retest correlations for the interview interaction variables were then recomputed for each of two subgroups: the ten individuals with the highest difference scores (High Change Index) and the ten with the lowest difference scores (Low Change Index). These recomputed correlations are shown in Table 23.

Comparison of Tables 21 and 22 with Table 23 clearly reveals that, for six of the variables shown, half the patients (Low Change Index group) had almost identical interviewee interaction patterns 8 months later. This can be seen from the unusually high coefficients of correlation (.91 to .98). The unusual stability in interviewee performance thus demonstrated had been masked when the two subgroups had been considered as one population (as in Tables 21 and 22).

It is evident also from Table 23 that the subgroup with a high change index showed practically no stability on five of

the six variables which had unusually high stability after 8 months in the group with a low change index (Pt.'s Units, Action, Tempo, Activity, and Dr.'s Units), and these attained coefficients of stability were lower than in any of our other test-retest samples, described earlier. That is, inspection of the data for the Low Change Index group shows that the consistency (stability) in their retest interview performance was not equally high for all the interview variables. Thus, for this group, the retest stability coefficients were small and statistically not significant for the following variables: Pt.'s Silence, Pt.'s Adjustment, Dr.'s Adjustment, Pt.'s Initiative, and Pt.'s Dominance. The first four of these variables are expressions of silence behavior during the interview. Since a recently completed and unpublished factor-analysis of the interaction scores of 60 patients had shown that interview silence and interview action behavior are important and independent factors for understanding patient-doctor interview interaction, we were not surprised that our first crude change index (derived from Pt.'s Action and Units)

TABLE 23

TEST-RETEST CORRELATIONS FOR INTERVIEW INTERACTION VARIABLES
AFTER EIGHT MONTHS FOR TWO SUBGROUPS RANK-ORDERED
ON PT.'S UNITS AND PT.'S ACTION

	<i>High Change Index†</i>		<i>Low Change Index†</i>	
	<i>r</i>	<i>rho</i>	<i>r</i>	<i>rho</i>
1. Pt.'s Units	.19	.32	.93**	.97**
2. Pt.'s Action	.03	.22	.95**	.95**
3. Pt.'s Silence	.58	.55	.35	.25
4. Pt.'s Tempo	.02	.22	.95**	.98**
5. Pt.'s Activity	.05	.02	.95**	.94**
6. Pt.'s Adjustment	.42	.32	.42	.50
7. Dr.'s Adjustment	.51	.65*	.37	.31
8. Pt.'s Synchronization	.49	.59	.93**	.91**
9. Dr.'s Units	.24	.20	.95**	.92**
10. Pt.'s Initiative	.80**	.83**	.49	.52
11. Pt.'s Dominance	.60	.54	— .15	— .21

* Significant at the .05 level of probability.

** Significant at the .01 level of probability.

† N is 10 in each of these subgroups.

should increase stability on one set of variables while decreasing it on another.

We therefore next divided the sample into two *new* subgroups according to the magnitude of test-retest *difference scores* on one of the silence variables, Pt.'s Adjustment. The results are shown in Table 24. It is clear from Table 24 that, as with the change index derived from the Action and Unit variables, there existed a subgroup of 10 patients with unusually high coefficients of stability for the variables related to Silence. Stability on these variables had been masked also when the subgroups were considered as one sample (Tables 21, 22).

Since number of psychotherapy hours received in the 8 month interval by the 20 patients in this Massachusetts General Hospital sample was in no way under our control, we considered the possibility that a closer examination of this variable might throw light on the observations in Tables 21, 22, 23, and 24. Table 25 shows the results of this examination for the two subgroups whose index of change had been derived from the variables Pt.'s Units and Pt.'s Action. The results of

this "experiment of nature" indicate that the 10 patients with test-retest stability coefficients reaching as high as .91 to .98 after 8 months for 6 of the variables had had significantly fewer psychotherapy hours (2.3) than had the 10 patients with the correspondingly higher change index (7.2 psychotherapy hours). Thus, concomitants of the greater number of psychotherapy hours may have produced the considerable retest changes (and thus lower retest reliability) in the interview interaction pattern of the 10 patients in the High Change Index group. As is evident from Table 26, patient variables other than number of psychotherapy hours which we also examined did not differentiate the two groups.

Tables 27 and 28 are similar in character to Tables 25 and 26 except that the comparisons are for the two new subgroups, rank-ordered on the variable, Pt.'s Adjustment. Neither number of psychotherapy hours (Table 27), nor the other variables shown in Table 28 differentiated the subgroup with the high change index from the subgroup with the low change index. Thus, if number of

TABLE 24

TEST-RETEST CORRELATIONS FOR INTERVIEW INTERACTION VARIABLES
AFTER EIGHT MONTHS FOR TWO SUBGROUPS RANK-ORDERED
ON PT.'S ADJUSTMENT

	High Change Index (N=10)		Low Change Index (N=10)	
	<i>r</i>	<i>rho</i>	<i>r</i>	<i>rho</i>
1. Pt.'s Units	.77**	.75†	.61	.59
2. Pt.'s Action	.57	.86**	.61	.62
3. Pt.'s Silence	.28	.29	.93**	.95**
4. Pt.'s Tempo	.51	.76†	.61	.67*
5. Pt.'s Activity	.61	.85**	.62	.57
6. Pt.'s Adjustment	.18	.22	.98**	.94**
7. Dr.'s Adjustment	.26	.32	.66*	.72*
8. Pt.'s Synchronization	.75†	.83**	.73†	.73†
9. Dr.'s Units	.85**	.93**	.60	.51
10. Pt.'s Initiative	.37	.47	.87**	.89**
11. Pt.'s Dominance	.34	.26	.27	.09

* Significant at the .05 level of probability.

† Significant at the .02 level of probability.

** Significant at the .01 level of probability.

TABLE 25

NUMBER OF PSYCHOTHERAPY HOURS BETWEEN TEST AND RETEST FOR
TWO SUBGROUPS WHOSE INDEX OF CHANGE WAS DERIVED
FROM PT.'S UNITS AND PT.'S ACTION

	<i>Hours of Psychotherapy Received</i>		<i>F-test</i>	<i>p</i>
	<i>High Change Index†</i>	<i>Low Change Index†</i>		
Mean	7.2	2.3	9.86	.01
Median	6.5	1.8		
SD	4.4	2.2	2.05	NS
Range	0 to 13	0 to 7		

† N is 10 in each of these subgroups.

TABLE 26

ORGANISMIC AND EXPERIENTIAL VARIABLES OF TWO SUBGROUPS
RANK-ORDERED ON PT.'S UNITS AND PT.'S ACTION

	<i>High Change Index†</i>	<i>Low Change Index†</i>	<i>F-test</i>	<i>p</i>
Age (Years)				
Mean	40.2	44.8	.40	NS
Median	35.5	46.0		
Range	16 - 80	15 - 73		
I.Q.				
Mean	101.6	96.2	.74	NS
Median	104.5	95.5		
Range	86 - 113	73 - 124		
Education (Years)				
Mean	10.6	9.2	2.07	NS
Median	10.5	9.0		
Range	6 - 15	7 - 11		
Interval between Interviews (Months)				
Mean	7.0	7.9	.98	NS
Median	7.9	8.0		
Range	6 - 9	3 - 12		
Patient Source				
Inpatient	3	2		
Outpatient	7	8		
Diagnosis				
Depression	3	5		
Anxiety Neurosis	2	4		
Hysteria	2			
Schizoid Personality		1		
Psychopath	1			
Grief Reaction	1			
Possible Psychotic	1			

† N is 10 in each of these subgroups.

TABLE 27

NUMBER OF PSYCHOTHERAPY HOURS BETWEEN TEST AND RETEST
FOR TWO SUBGROUPS WHOSE INDEX OF CHANGE
WAS DERIVED FROM PT.'S ADJUSTMENT

	<i>Hours of Psychotherapy Received</i>		<i>F-test</i>	<i>p</i>
	<i>High Change Index†</i>	<i>Low Change Index†</i>		
Mean	5.7	3.8	1.01	NS
Median	4.5	2.2		
SD	4.7	3.7	1.26	NS
Range	0 to 13	0 to 12		

† N is 10 in each of these subgroups.

TABLE 28

ORGANISMIC AND EXPERIENTIAL VARIABLES OF TWO SUBGROUPS
RANK-ORDERED ON PT.'S ADJUSTMENT

	<i>High Change Index†</i>	<i>Low Change Index†</i>	<i>F-test</i>	<i>p</i>
Age (Years)				
Mean	42.6	42.4	.00	NS
Median	44.5	40.5		
Range	15 - 80	29 - 73		
I.Q.				
Mean	99.8	98.2	.06	NS
Median	100.5	101.0		
Range	83 - 115	73 - 124		
Education (Years)				
Mean	10.2	9.7	.21	NS
Median	10.0	10.25		
Range	6 - 15	7 - 12		
Interval between Interviews (Months)				
Mean	6.8	8.1	2.17	NS
Median	6.8	7.5		
Range	3 - 10	5 - 12		
Patient Source				
Inpatient	5	2		
Outpatient	5	8		
Diagnosis				
Depression	5	3		
Anxiety Neurosis	2	4		
Hysteria	2			
Schizoid Personality		1		
Psychopath		1		
Grief Reaction		1		
Possible Psychotic	1			

† N is 10 in each of these subgroups.

psychotherapy hours is a relevant variable for producing change in interviewee interaction behavior over a time interval of 8 months, it appears to have significance for the action and action-related variables and not for the Pt.'s Adjustment and other silence-related variables. At any rate, whether or not we have identified the pertinent variable, i.e., number of psychotherapy hours received, the results are of interest in suggesting further controlled experiments.

MODIFIABILITY IN INTERVIEWEE
INTERACTION BEHAVIOR IN RELATION TO
PLANNED MODIFICATIONS IN INTERVIEWER
BEHAVIOR

In an earlier section of this paper we reviewed some of the relevant studies dealing with the effect of interactees upon each other during interviews. In previous publications we have presented evidence that the intra-interview modifications in the interviewer's behavior (as he goes from period to period) affect markedly certain interviewee variables (26).

Tables 2 and 3 define the prescribed changes in the interviewer's behavior (the independent variable). Tables 7, 8, 9, and 10 indicate to what degree he is able,

in fact, to modify his behavior in the prescribed ways. If we examine the relationship between the silence behavior of the patient as the interviewer modifies his own silence behavior from one subperiod to another, we obtain the results shown in Table 29. These data are the averages for the group of 20 patients from the Massachusetts General Hospital sample previously described.

Study of the data of the first interview for these 20 patients shows that the patient's silence behavior varies significantly from period to period, with the longest average duration of patient silence appearing in Period 2. Table 29 shows further that the difference among the means of the five subperiod silences is statistically significant (.001 level) as was determined both by analysis of variance (F-test) and non-parametric statistics using Friedman's Sign-test method (42, pp. 166-172). Analysis of differences between pairs of means is shown in Table 30. Here it is seen that the durations of patient silence in the two stress periods (2 and 4) differ significantly from the durations in all 3 remaining free periods (1, 3, and 5). The observations were clearly reproducible, as shown by the

TABLE 29

MODIFIABILITY OF INTERVIEWEE: SILENCE: MEANS, STANDARD DEVIATIONS
AND SIGNIFICANCE LEVELS FOR PATIENT SILENCE VARIABLE
ON TEST AND RETEST (MGH SAMPLE)

<i>Period</i>	<i>First Interview</i>		<i>Second Interview</i>	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
I	8.59†	1.84	7.94	1.74
II	13.92	6.02	14.24	6.57
III	9.04	2.11	8.58	2.58
IV	6.14	1.75	5.85	1.55
V	7.78	1.61	7.89	1.75
F-test	17.84***		19.97***	
Sign-test	37.47***		35.23***	

† All values are in hundredths of a minute. To convert to seconds, multiply the value shown by 0.6.

*** Significant at the .001 level of probability.

data under the second interview column (in Tables 29 and 30), and which were obtained during retest interview after a median of 8 months. Both sets of data show that the patients' longest average duration of silence is in Period 2, and shortest in Period 4. From the rules given in Table 3 it will be remembered that the interviewer's silence (prescribed

as 15 seconds and, as shown in Table 9, actually averages 16 seconds), is longest in Period 2, and, since he interrupts is, in a sense, shortest in Period 4. Since patients' longest silences (Period 2) occur in relation to the interviewer's longest silences (Period 2) one interpretation of these relationships which has occurred to us is that they represent a type of complex (operant) conditioning.

Evidence for a similar kind of operant conditioning relationship, between interviewer behavior and interviewee average duration of action, is shown in Table 31. The interviewee's actions are shortest in Period 4, when the interviewer is interrupting him, and next shortest in Period 2, when the interviewer, by failing to respond 12 times, has reduced his own actions to zero. This modifiability in the average duration of the interviewee's actions is also statistically significant as shown by the significant F and Sign-tests at the bottom of Table 31. Analysis of differences between pairs of means is shown in Table 32.

The results shown in Tables 29, 30, 31 and 32 are group trends (period by period averages for all 20 patients). Analysis of the degree to which *individual patients*

TABLE 30

MODIFIABILITY OF INTERVIEWEE SILENCE IN FIRST AND SECOND INTERVIEW DATA: PROBABILITY LEVELS OF t-TESTS OF DIFFERENCES BETWEEN PAIRS OF MEANS†

Periods	Periods (First Interview)				
	1	2	3	4	5
1		.01		.001	
2	.01		.01	.001	.001
3		.01		.001	.02
4	.001	.001	.001		.01
5		.001	.02	.01	

Periods	Periods (Second Interview)				
	1	2	3	4	5
1		.001		.001	
2	.001		.01	.001	.001
3		.01		.001	
4	.001	.001	.001		.001
5		.001		.001	

† t-tests were computed for correlated means.

TABLE 31

MODIFIABILITY OF INTERVIEWEE ACTION: MEANS, STANDARD DEVIATIONS AND SIGNIFICANCE LEVELS FOR PATIENT ACTION VARIABLE ON TEST AND RETEST (MGH SAMPLE)

Period	First Interview		Second Interview	
	Mean	SD	Mean	SD
I	45.40†	49.63	43.60	31.54
II	34.90	24.91	27.75	10.25
III	38.70	32.95	57.40	85.11
IV	7.45	4.29	6.50	2.37
V	41.75	48.47	34.65	31.42
F-test	6.62**		5.20**	
Sign-test	42.53***		45.83***	

† All values are in hundredths of a minute. To convert to seconds, multiply the value shown by 0.6.

** Significant at the .01 level of probability

*** Significant at the .001 level of probability.

showed the group modifiability trend from period to period (shown in both Tables 29 and 31) revealed the following. For the action variable (Table 31), 20 out of 20 patients showed a drop in action from Period 3 to 4 and an increase again in Period 5. This finding was replicated for all 20 patients in their second interviews. The drop in action from Period 1 to 2, with increase in Period 3 was shown only by 8 patients in the first interview, and by 11 of the 20 in the second interview. Thus there were 12 deviates in the first interview from this "drop in Period 2, increase in Period 3" aspect of the interview, and 9 deviates in the second interview. Five of the 12 deviates in the first interview were included among the 9 deviates of the second interview, thus showing a "stability" in their own deviant behavior from the group trend.

For the silence variable, analysis by individuals revealed that 14 out of the 20 patients showed an increase in silence from Period 1 to 2 and a decrease in Period 3. For the retest interview 8 months later, 16 out of the 20 subjects

showed a similar pattern. Interestingly, 3 out of the 4 deviates in the second interview were included among the 6 deviates from the group trend in the first interview. Thus, these individuals also were consistent in their deviant behavior. Analysis of patient silence behavior in Period 4 showed that for the first interview 16 of the 20 patients showed a decrease in silence from Period 3 to 4, with a corresponding increase in Period 5; 4 did not. On retest the figures were 17 and 3 respectively. These 3 deviates were among the 4 deviates in the first interview.

It would appear from all of the results presented on the interviewee reliability question that: (1) interviewee interaction patterns are stable under identical stimulus conditions, unique for any given individual, and reliably measurable; (2) interviewee interaction patterns are predictably and reliably modifiable under conditions of planned changes in interviewer behavior; and (3) many of the deviates from the observed group trends are reliably so on retest. Thus the Interaction Chronograph method may well have possibilities for the description of important facets of any single individual's personality by methods of more detailed analysis than we have so far used, perhaps involving sequential analysis of one person's interactions unit by unit.

TABLE 32

MODIFIABILITY OF INTERVIEWEE ACTION IN
FIRST AND SECOND INTERVIEW DATA:
PROBABILITY LEVELS OF *t*-TESTS OF
DIFFERENCES BETWEEN PAIRS OF MEANS†

Periods	Periods (First Interview)				
	1	2	3	4	5
1				.01	
2				.001	
3				.001	
4	.01	.001	.001		.01
5				.01	

Periods	Periods (Second Interview)				
	1	2	3	4	5
1		.02		.001	
2	.02			.001	
3				.02	
4	.001	.001	.02		.001
5				.001	

VALIDITY OF THE METHOD OF THE PARTIALLY STANDARDIZED INTERVIEW

The results described to this point indicate that interviewee interaction patterns have a high degree of reliability on the one hand, and considerable modifiability on the other. Since the purpose of this paper was primarily to describe the potential usefulness of the Interaction Chronograph method as a means of studying possible changes in behavior associated with psychotherapy, or some other interpolated experimental condition,

our attempts to date to study the validity of personality descriptions in terms of interview interaction patterns will be only briefly presented. Our approaches to the problem of validity have included examination of the relations between: (a) interviewee interaction patterns and psychological test scores, and other selected organismic characteristics (27); (b) content-analysis scores derived from the verbatim recordings of the standardized interviews and their relationship to simultaneously measured interviewee interaction scores (29); (c) a factor analysis of the Interaction Chronograph scores, themselves (to be published); (d) differences in interviewee interaction scores of five different diagnostic populations, ranging from normal, through outpatient, to chronic, hospitalized schizophrenic status (25); (e) the relationship between interviewee interaction scores and ward behavior of the same subjects observed during a one-week period (to be published); and (f) a simultaneous description of the behavior of both the interviewer and the interviewees by the Interaction Chronograph and a second observational method, Bales' Interaction Process Analysis (18).

SUMMARY: IMPLICATIONS FOR THE USE OF THE INTERACTION CHRONOGRAPH METHOD IN STUDIES ON PSYCHOTHERAPY

1. The interviewee variables measured by the Interaction Chronograph represent descriptions of observable behavior and thus do not pose the well-known problems of second and third-order inference as do, for example, global clinical impressions or the data from Rorschach, TAT, and other such psychological test procedures.

2. The reliability of the interviewee variables (the dependent variables) is sufficiently high, even after 8 months, to have them serve as an adequate index of change. In the main, this is no doubt due to the fact that the behavior of the

interviewer, as independent variable, is both specifiable and reliable.

3. Preliminary research indicates interviewee interaction scores are not related to content of the interview in such a way as to nullify their usefulness in studies of psychotherapy.

4. In several validity studies, the data of which were not presented here, the relationship of the Interaction Chronograph variables to a variety of organismic variables (age, sex, socio-economic status, intelligence, anxiety level, diagnosis, etc.) has been defined. Thus these important variables, often overlooked, can be specified, varied systematically, or otherwise controlled in studies on psychotherapy.

5. Preliminary observations from an "experiment of nature" have suggested that change in some of the Interaction Chronograph interview variables is related to number of hours of psychotherapy experienced.

The statements listed above, taken together, seem to meet the essential conditions stated by Edwards and Cronbach (11) for research in psychotherapy (or, for that matter, research on the effects of drugs, psychosurgery, use of therapists of differing levels of experience, differing theoretical frames of reference, the effects of planned modifications in the interviewer's behavior analogous to operant conditioning procedures, etc.). It may well be possible to design experiments dealing with psychotherapy (and other experimenter manipulanda) in which a limited number of stimulus variables and organismic variables is defined, and specified interviewee as well as interviewer response variables are measured by use of this highly reliable method.

REFERENCES

1. Ash, P. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1949, 44, 272-276.
2. Chapple, E. D. Quantitative analysis of the interaction of individuals. *Proc. Nat. Acad. Sci.*, 1939, 25, 58-67.

3. Chapple, E. D. "Personality" differences as described by invariant properties of individuals in interaction. *Proc. Nat. Acad. Sci.*, 1940, 26, 1, 10-16.
4. Chapple, E. D. The Interaction Chronograph; its evolution and present application. *Personnel*, 1949, 25, 295-307.
5. Chapple, E. D. The standard experimental (stress) interview as used in Interaction Chronograph investigations. *Human Organiz.*, 1953, 12, 23-32.
6. Chapple, E. D. *The Interaction Chronograph Manual*. Noroton, Conn.: E. D Chapple Co., 1956.
7. Chapple, E. D., with Arensberg, C. M. Measuring human relations: An introduction to the study of the interaction of individuals. *Genet. Psychol. Monog.*, 1940, 22, 3-147.
8. Chapple, E. D., Chapple, M. F., & Repp, J. A. Behavioral definitions of personality and temperament characteristics. *Human Organiz.*, 1954, 13, 34-39.
9. Chapple, E. D., & Donald, G., Jr. A method for evaluating supervisory personnel. *Harvard Bus. Rev.*, 1946, 24, 197-214.
10. Chapple, E. D., & Lindemann, E. Clinical implications of measurements of interaction rates in psychiatric interviews. *Appl. Anthropol.*, 1942, 1, 1-11.
11. Edwards, A. L., & Cronbach, L. J. Experimental design for research in psychotherapy. *J. clin. Psychol.*, 1952, 8, 51-59.
12. Gleser, Goldine, Haddock, J., Starr, P., & Ulett, G. Psychiatric screening of flying personnel: Interrater agreement on the basis of psychiatric interviews. USAF Sch. of Aviation Med., Project No. 21-0202-0007, Rep. No. 10, Nov. 1954.
13. Goldman-Eisler, F. The measurement of time sequences in conversational behavior. *Brit. J. Psychol.*, 1951, 42, 355-362.
14. Goldman-Eisler, F. Individual differences between interviewers and their effect on interviewees' conversational behavior. *J. Ment. Sci.*, 1952, 98, 660-671.
15. Goldman-Eisler, F. A study of individual differences and of interaction in the behavior of some aspects of language in interviews. *J. Ment. Sci.*, 1954, 100, 177-197.
16. Goldman-Eisler, F. On the variability of the speed of talking and its relation to the length of utterances in conversations. *Brit. J. Psychol.*, 1954, 45, 94-107.
17. Greenspoon, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctoral dissertation, Univer. Indiana, 1950.
18. Hare, A. P., Waxler, Nancy, Saslow, G., & Matarazzo, J. D. Interaction process analysis in a standardized initial psychiatric interview. Submitted for publication.
19. Hunt, W. A., Wittson, C. L., & Hunt, Edna B. A theoretical and practical analysis of the diagnostic process. In P. H. Hoch & J. Zubin (Eds.), *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953, 53-65.
20. Kelly, E. L. In C. P. Stone (Ed.), *Annu. Rev. Psychol.*, Stanford, Calif.: Annual Reviews, Inc., 1954, 5, 281-310.
21. Krasner, L. A review of "verbal conditioning" studies with their implications for psychotherapy. Paper read at the Sixteenth Bay Area Research Conference, held at the VA Hospital, San Francisco, June 13, 1957.
22. MacKinnon, D. W. The structure of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944, 1, 3-48.
23. Matarazzo, J. D., Saslow, G., & Matarazzo, Ruth G. The Interaction Chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, 41, 347-367.
24. Matarazzo, J. D., Saslow, G., & Guze, S. B. Stability of interaction patterns during interviews: a replication. *J. consult. Psychol.*, 1956, 20, 267-274.
25. Matarazzo, J. D., Saslow, G., Matarazzo, Ruth G., & Phillips, Jeanne S. Differences in interview interaction patterns among five diagnostic groups. Paper read at Amer. Psychol. Ass., New York City, Sept., 1957.
26. Matarazzo, J. D., Saslow, G. Matarazzo, Ruth G., & Phillips, Jeanne S. Stability and modifiability of personality patterns during a standardized interview. In P. A. Hoch and J. Zubin (Eds.), *Psychopathology of communication*. New York: Grune & Stratton, 1958, pp. 98-125.

27. Matarazzo, Ruth G., Matarazzo, J. D., Saslow, G., & Phillips, Jeanne S. Psychological test and organismic correlates of interview interaction behavior. *J. abnorm. soc. Psychol.*, 1958, 56, 329-338.
28. McNemar, Q. *Psychological statistics*. New York: Wiley, 1955.
29. Phillips, Jeanne S. The relationship between two features of interview behavior comparing verbal content and verbal temporal patterns of interaction. Unpublished doctoral dissertation, Washington Univer., St. Louis, 1957.
30. Phillips, Jeanne S., Matarazzo, J. D., Matarazzo, Ruth G., & Saslow, G. Observer reliability of interaction patterns during interviews. *J. consult. Psychol.*, 1957, 21, 269-275.
31. Raines, G. N., & Rohrer, J. H. The operational matrix of psychiatric practice: I. Consistency and variability in interview impressions of different psychiatrists. *Amer. J. Psychiat.*, 1955, 11, 721-733.
32. Rogers, C. R. *Counseling and psychotherapy*. Boston: Houghton Mifflin, 1942.
33. Rogers, C. R., & Dymond, Rosalind F. *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.
34. Salzinger, K., & Pisoni, Stephanie, Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, 57, 84-90.
35. Sarason, S. B. *The clinical interaction. with special reference to the Rorschach*. New York: Harper, 1954.
36. Saslow, G. Psychotherapy. In C. P. Stone (Ed.), *Annu. Rev. Psychol.*, Stanford, Calif.: Annual Reviews, Inc., 1954, 5, 311-336.
37. Saslow, G., Goodrich, D. W., & Stein, M. Study of therapist behavior in diagnostic interviews by means of the Interaction Chronograph. *J. clin. Psychol.*, 1956, 12, 133-139.
38. Saslow, G., Matarazzo, J. D., & Guze, S. B. The stability of Interaction Chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, 1955, 19, 417-430.
39. Saslow, G., Matarazzo, J. D., Phillips, Jeanne S., & Matarazzo, Ruth G. Test-retest stability of interaction patterns during interviews conducted one week apart. *J. abnorm. soc. Psychol.*, 1957, 54, 295-302.
40. Saslow, G., & Peters, Ann deH. A follow-up study of "untreated" patients with various behavior disorders. *Psychiat. Quart.*, 1956, 30, 283-302.
41. Sidowski, J. B. Influence of awareness of reinforcement on verbal conditioning. *J. exp. Psychol.*, 1954, 48, 355-360.
42. Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Psychophysiological Approaches to the Evaluation of Psychotherapeutic Process and Outcome¹

JOHN I. LACEY, PH.D.

Unlike the other contributors to this conference, I have not been engaged directly in the experimental study of the therapeutic process. Psychophysiological studies of psychotherapeutic process and outcome, however, are slowly accumulating. I have been asked, therefore, to review this area of research, in the light of more general considerations of the psychophysiology of the autonomic nervous system.

The relationship of the autonomic nervous system to behavior in general, and to the problems of neurosis and psychosis, is a vast and controversial field. It would be impossible in this conference to review and critically discuss the material concerning the moot role of the autonomic nervous system in the etiology of disordered behavior, or concerning the use of autonomic measurements for prognosis. It would be equally impossible to consider the vast relevant field of psychosomatic medicine. Time limitations, considerations of personal competence and experience, as well as my understanding of the aims of this conference, lead me to focus attention on those investigations which utilize measurements of peripherally accessible functions such as heart rate, skin temperature, muscle potentials and skin resistance in the course of therapeutic or quasi-therapeutic interviews.

The potential contribution of such psychophysiological studies of the psychotherapeutic process is alleged by many to be great. The reasons stated or implied are obvious. Such measures as skin resistance, heart rate, blood pressure, blood flow, skin temperature, blood-oxygen saturation, gastric motility, pupillary diameter, muscle tension, and other variables have been demonstrated to be remarkably sensitive and responsive measures in a variety of "emotional" states. Conflict, threat and frustration; anxiety, anger, and fear; startle and pain; embarrassment; pleasant and unpleasant stimuli—all these produce autonomic changes. The "threat" to the organism can be real or imagined, present or recalled or anticipated, social or physical, verbalizable or totally inaccessible to verbalization at the time. In all these situations many physiological changes occur. The interview situation itself is rich in observable somatic changes. Indeed, in predisposed individuals, painful and even dangerous somatic changes—such as headache (16, 49, 50, 51), backache (33), episodes of Raynaud's disease (55), production of blood, bile and excessive hydrochloric acid in the stomach (56)—can be precipitated by the discussion of conflictful and psychologically threatening material.

Manifold physiological changes occur in a bewildering variety of situations, and are seen even when it is difficult to attach a traditional emotional label—such as anger, or hostility, or self-contempt—to the state of the organism. Simple physical stimuli that are not painful, or startling, or intense can produce the same

1. Preparation of this paper was supported in part by a grant, M-623, from the National Institute of Mental Health, United States Public Health Service. Most of the published papers from our laboratory that are reviewed in this paper, and all of the unpublished material, derive from investigations supported in part by this grant.

physiological changes, especially if they are novel intrusions into the environment. Hence, a simplification, derived either from Cannon-like conceptions of the "energy mobilizing" functions of autonomic (especially sympathetic) activity, or from electrophysiological concepts of cortical "arousal," has been introduced by several modern theorists (11, 27, 28, 44, 46, 75). These physiological changes are considered to be *indicators* of the state of "arousal," "activation," "energy level," "behavioral intensity," or general "tension" of the organism. A uni-dimensional continuum is postulated, ranging from the deepest of sleep to extremely excited panic states. The "activation level" is proposed as a measure of the motivational or emotional significance of a given situation to the individual, and as a measure related (albeit in complicated fashion) to various characteristics of behavior (28, 46, 48, 67).

In a word, autonomic and skeletal-motor functions are commonly utilized as a metering or indicating function, measuring intensity of affect and intensity of arousal. This is a *substitutive* use of physiological measurements, analogous in some ways to the lie-detection procedure. The physiological measures allegedly are more objective, more sensitive, easier to measure, and clearer in meaning than other, more directly psychological, measures and observations.

This "affect-meter" (or "arousal-meter") is thought by many to be an extremely useful tool in a breath-taking variety of undertakings. Thus, it is hoped that physiological measurements might enable us to evaluate subtle and otherwise not-easily-detectable changes in affect or arousal in the course of the therapeutic interview, and thereby help elucidate and quantify the affective interchange between therapist and patient. It is held to be at least possible that by the observation of autonomic and motor changes one could detect conflict areas

and their interrelationships somewhat in advance of their clear emergence in the verbal interaction of the interview. It is hoped that, by measuring the somatic impact of different psychotherapeutic techniques, and of different therapists, some clues might be secured to the differential effectiveness of therapists and techniques. It is hoped that autonomic and motor measurements can provide an objective means of substantiating the changes which purportedly occur during psychotherapy.

Laboratory and clinical investigations both provide data, remarkably congruent, in support of at least some of these hopes and expectations. There are enormous difficulties, however, and there are at least some suggestions to make of somewhat different approaches to the use of psychophysiology in the study of psychotherapy. Let me now turn to a review of the relevant findings.

AUTONOMIC AND SKELETAL-MOTOR CHANGES: INDICANT FUNCTIONS

Physiological changes as a quantitative indicant of focal conflict-areas. Physiological changes obviously are not all-or-none. Their magnitudes have a lawful relation to physical stimulus-characteristics and to analogous psychological characteristics of experience. This has been shown to be true in innumerable investigations, and it would be out of place to review them critically here. As important examples, however, we might mention a few, especially those which will serve to introduce some critical remarks I will make later in this paper. Hovland and Riesen (34) showed that the amplitudes of galvanic skin response and of vasoconstriction in the finger were linearly and positively related to the intensity of sound, and to the intensity of electric shock. Davis, Buchwald, and Frankmann (19) showed that a wide variety of physiological variables exhibited increasing magnitude of response

as the intensity of sound stimuli increased. These investigators measured muscle potentials, skin resistance, respiratory rate, heart rate, finger volume, chin volume, and related measures. Most of these responses showed significant relationships to sound intensity. Doust and Schneider (25) measured capillary blood-oxygen saturation with an ingenious and innocuous technique, and found that simply asking subjects to look at undifferentiated patches of color produced significant anoxemia, the degree of anoxemia being related to the color. In increasing order of anoxemic effect, the colors were blue, green, orange, yellow, purple, and red, a sequence roughly related to the rank order of color preferences for the subjects in the group studied. Furer and Hardy (30) measured skin resistance responses day-after-day to graded intensities of pain in 4 subjects, until the responses were almost completely adapted, and found consistent correlations between the magnitude of galvanic response and the physical intensity of the pain-stimulus, complicated, however, by the "threat-content" of the pain-stimulus to the subject.

As McCurdy (54) trenchantly points out in reviewing the Hovland and Riesen study, such variation in physical stimulus parameters are effective in producing variations of experience, and the results could just as well be displayed as increases in the magnitude of physiological responses with increases in perceived and judged intensity. Whatever the scientific and philosophical merits of this point of view, it is abundantly clear that physiological responses do increase as judged intensity increases. These judgments, moreover, need not be "simple" judgments of the intensity of a pain, the loudness of a sound, or the brightness of a color. They can be simplified reports of the "intensity" of introspectively complex material. McCurdy, in his vigorous and lively review of most of the data

concerning the relationship of skin resistance changes to such judgments, writes: "From the earliest period of PGR experimentation it has been asserted that magnitude of galvanometric deflection is intimately associated with the subject's estimate of the intensity of his experience—whether emotional, affective, conational or otherwise. This clear statement has been beclouded by sundry differences of opinion as to the exact quality of experience usually involved; but so far as the actual evidence goes, there has apparently never been an instance in the literature where the central issue has been in doubt."

McCurdy has performed a most valuable service, by re-analyzing the data, wherever he could in reviewing the literature, and computing contingency or product-moment correlations between galvanometric deflection-magnitude and judged "intensity of experience." The experiments reviewed were richly diversified, of course, including various stimuli and tasks, various judgmental and introspectional arrangements, various subjects, various devices for measuring PGR, and various kinds of arithmetic in arriving at quantification of magnitude of response. His coefficients are worth repeating here. They are, in order of increasing magnitude: .45, .45, .53, .58, .59, .59, .66, .67, .72, .78, .80, .81, .82, .86, .87, .88, .93, .94, and 1.00. This is a startling array of correlations. They are all positive, most are high, and some of them are extraordinarily high. The median value is +.75. The "intensity of affective experience" is certainly correlated with magnitude of galvanometric deflection.

McCurdy also emphasizes the striking reliability of results relating galvanometric deflection to specific words: "The 'emotional intensity values' found galvanometrically by Whately Smith for a word list applied to subjects in England was confirmed for a number of words by Jones and Wechsler when they were

applied several years later to subjects in America." Stimulated by this suggestion, my wife and I had small groups of elementary psychophysiology students administer the first ten and last ten of the Whately Smith word list to volunteer subjects. Each student experimenter ran one subject, taking care to impose each word upon the same pre-stimulus level of skin resistance for that subject, and calling for the first association that occurred to the subject. For three different groups of 7, 12, and 9 individuals, we secured rank-order correlations between our galvanometrically ranked words and Whately Smith's galvanometrically ranked words of .89, .87, and .94. In accord with McCurdy, we are certainly willing to entertain the suggestion that the differential magnitude of galvanometric deflection to words is one of the most reliable phenomena in psychology today!

It is not known in anywhere near this detail whether similar relationships hold for other somatic responses to provocative symbolic stimuli. Rowland (60), in a seemingly forgotten study, deeply hypnotized four subjects and on each of 10 days suggested, by means of a typed phrase on a card, a series of 12 situations which varied in their "exciting character." The "exciting character" of the stimuli had been established previously by the ratings of 29 judges of a total of 88 behavior situations. The typed phrase was meant to recall to the subject the complete description of the situation, which had been previously memorized under hypnosis. A variety of somatic effects were measured: pulse rate, galvanic skin response, and six different respiratory measures. The results show that physiological levels and responses increase roughly as a function of the scale value or "exciting character" of the stimulus. Comparing the three stimuli lowest, and the three stimuli highest, in "exciting character," there is no overlap

in the averages for any of the physiological variables.

Stennett (67), in Malmo's laboratory, has shown that palmar conductance and muscle tension systematically increase as motivational conditions are varied from very low to very high levels, and Malmo and Davis (48) have secured other data suggesting that in addition heart rate, blood pressure, and respiration are similarly responsive to changes in arousal conditions.

Doust and Schneider (26) have extended their technique of measuring capillary blood-oxygen saturation to a study of the effects of word and picture stimuli. They had 3 alternative sets of 15 stimuli, 7 of which were "neutral" in affect and 8 of which were potentially "traumatic or stressful." In each of these stimulus-sets, ten were words taken from Rapaport's Word Association list, and the other five stimuli were four TAT and MAPS cards, and one Rorschach card. Large N's, and many different normal and psychiatric samples, were used. The "neutral" or "non-traumatic" stimuli produced no significant changes in capillary blood-oxygen saturation from resting levels. All the potentially stressful or traumatic stimuli, however, produced highly significant depressions in blood-oxygen values for all groups studied except depressives and constitutional schizophrenics.

In all these results there is the suggestion of a startling universality of result. Words such as kiss, love, wife, sex, seem to almost universally produce more response than words such as dance, afraid, or hunger. Is the technique sensitive enough to yield quantitative differences of autonomic activation that are related to the conflict areas within an individual? The answer seems to be "Yes." There are a few experimental studies, and the question has been studied in considerable detail in the confines of the therapeutic or quasi-therapeutic interview.

In the just mentioned study by Doust and Schneider on the production of relative anoxemia by words and pictures, a few case studies are given to illustrate the detection of areas of focal emotional conflict by the measurement of blood-oxygen saturation. While these are not presented in sufficient numbers to enable evaluation of the generality of the results, these three case studies do suggest a clear-cut differentiation between personally significant and non-significant stimuli. As an example. "Mr. T. W., aged 29, came to the hospital complaining of feelings of tension and anxiety in the evenings when he returned home from his work. Throughout the day he was a fit and happy man; only when his work was finished did his symptoms begin to trouble him." Shortly after his hospitalization, this patient was subjected to the measurement of blood-oxygen saturation responses to a series of word stimuli. Markedly increased anoxemic episodes were provoked by two words, "wife" and "sex." The production of anoxemia in the subject by these two words would certainly be an extremely early clue to the therapist as to the nature of this individual's complaint—if his clinical intuition had not already told him what to look for.

In a similar vein, Sines (65) has reported a study in which 20 patients were subjected to 4 pictures. One of these pictures was a neutral TAT card, one a picture, especially drawn for the study, related to hostility, one a TAT card designed to evoke the theme of passive dependency, and one an Esquire picture with a sexual theme. These 20 patients were divided into 3 groups; apparently the differentiation was quite clear. One group consisted of patients for whom it had been already established that problems of handling hostility evoked clinically apparent anxiety; for a second group, anxiety derived from dependency conflicts; and for a third group the anxiety

was centered around sexual problems. Sines measured respiration rate, heart rate, and GSR responses and averaged the responses to get a measure of overall autonomic nervous system activation by these pictures. Analysis of variance showed that hostile patients reacted more to the hostility picture than to the other pictures; dependent patients reacted more to the dependency stimulus than to the sex or hostility pictures; sexually conflicted patients reacted more to the sex picture than to the other pictures. The interaction between groups and stimuli was significant at the 1% level. It is not clear whether there was some selection of data in the Sines study. It is stated that other stimuli were used, but that the reported results are for the picture stimuli that yielded positive results. Depending upon the variety and number of other stimuli that were used, we may question whether there was not capitalization on chance findings. However, these results are consistent with all other results.

In therapeutic or quasi-therapeutic interviews, or in especially designed experiments, quantitative differentiation of motor and autonomic response related to the evocation of conflict-laden material has been reported many times. There is a complete unanimity of result that the flow of verbal interaction between therapist and patient is punctuated by autonomic and motor upheavals that are related in time, *moment-by-moment*, to the discussion of personally significant material. The shifts in content of the therapeutic interchange, and shifts in moment-to-moment interactions between patient and therapist are reflected in parallel somatic events. Mittelman and Wolff (55, 56, 57) have given many protocols illustrative of this phenomenon, using finger temperature and measures of gastric function. Wolf, Wolff and their colleagues (31, 32, 74) have also contributed dramatic findings on gastro-

intestinal and nasal reactions to the discussion of meaningful material. Many of these clinical observations strongly suggest that the magnitude of somatic change observed in the highly variable and uncontrolled (so far as physiological measurement is concerned) interview situation is as differentially related to the depth and complexity and significance of the verbal material as in more tightly controlled experimental studies. For example, Mittelman and Wolff (55) report finger temperature changes for a subject who read the pamphlet, *Sudden Death*, which presents in harrowing and gruesome detail the results of automobile accidents. The subject showed a drop in finger temperature of only 1.8°C in 20 minutes. This, while a significant fall, is not as dramatic and sustained as other falls in this same subject. In reading an account of a dedicatory address by President Roosevelt, this subject exhibited a drop of 5.3°C. This apparently bland material, however, was evocative of deeply conflictual material, because Roosevelt was a source of strong disagreement between herself and her brothers. The apparently innocuous reading material provoked mixed feelings of anxiety, insecurity, anger, and resentment.

Similarly, Holmes and Wolff (33) have presented examples of sustained increase of muscle potentials in individuals susceptible to backache, arising during the discussion of threatening life situations. Malmo and his collaborators in a number of papers have presented clear evidence of the rise in muscle tension recorded from various muscle groups as individuals discuss emotionally significant material (16, 50, 51, 52, 63).

In pre-disposed patients, indeed, discussion of personally significant topics are capable of producing symptoms even when no symptoms are present at the onset of the interview. Davis and Malmo (16) report such a case where, in the

course of the stress engendered by interview, frontalis muscle tension rose to high and stable levels and precipitated headache, and the other already-cited papers from this group report additional cases of similar import. Holmes and Wolff (33) report several instances where backaches were precipitated by the specific stress of interview. Mittelman and Wolff (55) report the precipitation of sustained drop in finger temperature accompanied by pallor and pain in a patient with Raynaud's disease. In another paper, Mittelman and Wolff (56) report the precipitation of pain and the production of bile and blood appearing in stomach secretion as a result of the stress of the interview.

In an extraordinarily interesting and detailed series of observations (52, 63), Malmo, Shagass, and their collaborators, found the appearance of high muscle tension in the forearms associated with discussion of hostility themes, but in the legs when conflictual sexual material is discussed. Their minute-by-minute plots of muscle tension and of the content of the interview, and their detailed statistical analyses, constitute a convincing although small series of cases in which such specific associations were produced. They contend, and produce some evidence in support of their contention, that it is not correct to think of these findings as instances of ideomotor action. That is to say, there was evidence that increased muscle tension was not a covert muscular accompaniment of imagined or anticipated action. They interpret the increased tension as a motor manifestation of conflicting neural impulses in the central nervous system.

Dittes (23), in a case study of one patient, has shown that the exploration by the patient of sexual matters is accompanied by GSRs, which gradually disappear as therapy progresses, but which are reinstated by real life experiences, or

by dreams which increase the sexual conflict. Dittes also shows that the decrease in skin resistance response to sexual discussion is not attributable to a generalized decrease of somatic reactivity but is due to specific extinction of somatic response to exploration of this conflict area. This is shown by the fact that while sequence of hours did not correlate significantly with number of GSRs in the hours, the correlation between sequence of hours and the percentages of sexually-related verbalizations that were accompanied by GSRs was —.85.

This last bit of analysis, it should be mentioned here, is most important. Adaptation is a very common and striking phenomenon in all autonomic responses. It would be expected that autonomic activation, as a generalized response to repetitive exposure to the originally uncertain and anxiety-producing interview situation, would decrease. Just such a result has been reported by Martin (53), who studied skin resistance as a function of four successive experimental interviews.

Physiological changes as a more general indicant. In addition to the observations, reviewed above, of striking moment-to-moment shifts in somatic activity correlatable with discussion of specific conflict-laden content, there is a scattering of more general information relating the overall tone and content of the therapeutic hour to average levels of physiological activity for that hour, or relating moment-to-moment shifts in physiological activity to the generally affective nature of the therapeutic interaction.

Coleman, Greenblatt and Solomon report on observations of heart rate in a sequence of 44 interviews with one patient (12). This paper is remarkable for two things. It demonstrates an exquisite sensitivity of heart rate in indicating momentary shifts in affect, and it shows concomitant effects on the thera-

pist, paralleling the effects on the patient (of which, more later). These authors used either *single* cardiac periods, or *only two* successive cardiac periods, and correlated these constantly shifting and momentarily changing periods (translated into beats per minute for the single or double R-R interval) with judgments of the momentary affective state of the patient. These judgments were recorded by an independent observer of the therapeutic interaction, and were categorized as "depression," "anxiety," "extra-punitive hostility" and "intra-punitive hostility." The bases for the judgments are stated to be such things as hand gestures, the droop of mouth, the rate of speech, pitch of voice, and steadiness of voice. These were judgments of apparently fleeting and quite momentary episodes. While no general statement of the temporal unit of observation is made, their illustration shows an incident of "depression" occurring between two ventricular systoles! These momentary episodes were significantly correlated with equally momentary shifts of cardiac rate. The patient showed the highest heart rate during moments of "anxiety," the lowest heart rate in "depression" and intermediate heart rates in "extra-punitive hostility" and "intra-punitive hostility."

Anderson has published a between-interviews content analysis of 10 interviews with a single patient (2). A combined index of heart rate and heart rate variability, abbreviated IHV, derived from a continuous electrocardiographic record taken throughout each interview, was correlated with specific aspects of the content of the therapeutic hour. In interviews in which the client expressed large amounts of affect, IHV rose, indicating high stable levels of heart rate. For Anderson's case, IHV revealed the greatest "physiological tension" when the patient focused critically upon himself and his present behavior. The rated in-

tensity of affect was also significantly related to IHV.

In a similar fashion, DiMascio, Boyd, and Greenblatt (21) report a case study of one patient through twelve interviews. Electrocardiograms and skin temperatures of the patient were measured, and the therapist's affective behavior was also taken. The tape recorded interviews were analyzed broadly by means of Bales Interaction Process Analysis categories, and the distribution of the occurrence of affective and non-affective categories was ascertained by computing the percentage of incidents in each interview in which each of 5 social interaction measures appeared. The "neutral" and "disagreement" social interaction units were not significantly related to the physiological measures, but the incidence of positive and negative interactions (excluding "disagreement"), however, were. In interviews during which the patient showed much "tension release" his heart rate tended to be slow ($r = -.58$); in interviews in which the patient showed much "tension," his heart tended to be fast ($r = +.69$); the "neutral" interviews, regarded as midway in the "tension continuum," were insignificantly correlated with heart rate ($r = +.11$). Other relations of similar import are reported in the paper. In general, only the affective areas were significantly correlated with physiological measures.

In an earlier communication by DiMascio, Boyd, Greenblatt and Solomon (22), results are presented and interpreted to mean that "positive affect" (as defined by the Bales system of categorization) is associated with higher heart rates and more "sweating" than is "negative affect." This is based upon one interview with a schizophrenic patient. A fragment of data is also presented, however, for three other subjects, showing consistent association of "positive affect"

with higher heart rates, and of "negative affect" with lower heart rates.

Physiological changes and therapeutic course. So far, the investigations reported have been at a microscopic level, so to speak. The complex course of the therapeutic hour, the ebb and flow of discussion of significant and insignificant material, the complex skein of affect—all are recorded with apparently exquisite sensitivity and faithfulness in what R. C. Davis has called the somatic "sea of response." For the most part, we have dealt with relatively momentary displacements of continuously recorded physiological functions within single interviews, and the time-unit of observation has been small. There is another group of investigations where the time-unit of observation is much grosser, and physiological evidence is gathered discontinuously. The aim of these investigations is to depict, in the simplest of terms, the course of entire sequences of therapeutic interviews; to trace the rise and fall of tension over long sweeps of time. Only Dittes (23), in the study already cited, has combined the microscopic view (momentary changes induced by the discussion of sexual matters) with this more macroscopic view.

Mowrer and his colleagues (58) pioneered the application of measurements of finger-tip sweating to the study of grosser long-term physiological changes in therapy. The finger-tips, painted with ferric chloride, are pressed for three minutes on paper impregnated with tannic acid. The perspiration on the finger-tip carries the ferric chloride into solution with the tannic acid, and a dark stain appears. The density of the stain is proportional to the amount of perspiration, and may be measured by reflectometric or densitometric techniques. By taking such finger prints before and after each interview, a study of the total impact of a single interview was made, and

long-term changes in the course of many interviews were also studied.²

Mowrer *et al.* present detailed protocols and analyses of the relationship of stain-density, taken before and after each interview, to self-ratings by the patient of subjective tension. Plots are given for individual patients over many hours of client-centered therapy. In general, palmar sweating and self-tension ratings were obviously related. The changes in successive determinations of stain density could be correlated with general events in therapy. "Successful" cases were, in general, characterized by an inverted V pattern, in which tension rose over a series of interviews, and then fell to lower values than were found initially. On this point, the authors comment that while the inverted V pattern is not universal, it is the most nearly typical pattern, and, they claim "... it makes good sense theoretically since the first phase of therapy may be thought of as a period in which the neurotic organization of the patient's old "self-system" undergoes dissolution, and the second phase is one in which a new and more workable personality pattern is being developed, one which, in the end, enables the patient to achieve a higher degree of integration and a lower degree of tension than was

possible with the old pattern."³ Patients who left therapy prematurely showed long-term rises in palmar sweating. The maintenance of high stain density was also correlatable with punishing experiences in therapy. The ratio of the incidence of "discomfort" to "relief" words, in the transcribed recordings of the interview, were significantly correlated (about .55) with after-interview stain densities. These preliminary data seem most promising. The detailed case analyses, and the simple and direct statistical analyses, suggest that such simply obtained and easily quantified measures yield information of value in depicting the longitudinal course of therapy.

Surprisingly, in the almost five years that have elapsed since these data were published only one other study using this technique has been published, and deals only in small part with the study of psychotherapy. Bixenstine (4) secured daily finger prints from a graduate student and his wife. Every eighth day, these obliging hands yielded finger prints every hour, thus providing records of diurnal and weekly variation. The procedure was continued for six months until the summer vacation, and was then resumed for two weeks in the Fall. An attempt was made to relate the events of the perspiration curve to the flow of events in the life of this couple, and many suggestive coincidences were found. For the purposes of this paper, it may be noted that the onset of therapy for the graduate student was accompanied by a sharp rise to high sweating levels, which gradually diminished. The student also made prints before and after each therapeutic interview. Four of the interviews were very "punishing" for Mr. M., as revealed by marked increases of

2. Wenger and Gilchrist (73), it should be noted, found a correlation of only .31 between measurements of palmar conductance and this measure of palmar sweating. It is true, as Mowrer *et al.* point out (58), that the measurements of palmar sweating and palmar conductance were taken from different skin areas, with the subjects seated for the fingerprint procedure and standing for the palmar conductance measurement. Finger-print stains, moreover, were taken without requiring constant pressure on the paper. Despite these differences, this is a startlingly low correlation. It indicates, possibly, the pre-secretory nature of skin resistance changes. It may also indicate unsuspected difficulties in the stain technique.

3. Anderson (2), it may be noted here, found an inverted V in his case study, using a measure combining heart rate and heart rate variability into a single measure.

post-interview sweating over pre-interview sweating. Due note is made in the report that concomitant with this evidence of the distressing effect of the therapeutic hour, Mr. M. suddenly increased the interval between his appointments. During this period, Mr. M's record of finger sweating, obtained in the daily sampling, showed a dramatic and reliable drop to lower levels than ever obtained before. This decrease in sweating was maintained in the two weeks of recording in the following Fall. "In other words, it appears that during the time Mr. M. was undergoing therapy, something happened to his level of PS production which was still apparent some time after withdrawing from therapy." The implication, of course, is that Mr. M. was no longer "tense"; his therapy was "successful."

Palmar sweating is not, of course, the only (or even the best) physiological measure capable of reflecting long term changes in "tension," or "mood," or of reflecting the apparent "success" of therapy. There are, however, surprisingly few observations. Alexander (1) studied a hypertensive patient in analysis. Pre-interview and post-interview blood pressure determinations were made. Their values were shown to reflect the over-all content of the interview. They were higher in interviews classified as "very disturbed," less in "somewhat disturbed" interviews, and least in "calm" interviews. Towards the end of the 201 sessions reported, this hypertensive patient's blood pressure slowly dropped and became less variable.

Malmo and his associates (51, 63) have reported several cases where muscle tension seemed to be a sensitive indicator of clinical status "assessed in terms of general tension or anxiety." The clearest case is reported by Shagass and Malmo (63). They measured the mean tension in five separate muscle groups for a series of interviews with one patient. These

were correlated with routine ward notes of behavior and mood, which had been made independently by ward nurses who did not know their notes were to be used in this way. A striking concordance between the two measures was found. Lower muscle tension always accompanied cheerful mood; higher tension, depressed mood. The mean muscle tension also reflected the general course of therapy, and steadily diminished as the patient improved. That this was not due to general habituation to the recording situation was shown by sharp increases when the patient was informed of impending discharge from the hospital. After working through this problem in therapy, the patient once more reverted to her improved clinical status, and exhibited lower muscle tension than ever before.

The behavior of the therapist as an influence on the patient's physiology. That the therapist's behavior affects the patient is practically a postulate of psychotherapy. Physiological measurements apparently again are sensitive and responsive indicants of the affective interchange between patient and therapist.

DiMascio, Boyd, Greenblatt and Solomon report, unfortunately in inadequate detail, the heart rates of three patients interviewed by three separate interviewers (22). Regardless of the nature of the affective interaction (classified by the Bales system as "positive affect," "negative affect," or "neutral" social interactions), and regardless of the order in which the therapist interviewed the subjects, one psychiatrist consistently produced a lower heart rate in all three subjects. No attempt was made in this preliminary communication to account for this effect.

Mittelman and Wolff (56) report another fragment of information. "A shy continuously tense woman was not able to relax and had a low finger temperature (22°C) even under "control"

conditions in the presence of observer A. In the presence of observer B she could relax, with high finger temperature (ca. 35°C). After better adjustment, she could partly relax and her finger temperature reached a high level quickly in the presence of observer A in an overheated environment (28°C)."

Dittes, in a remarkably well-analyzed and interpreted paper, has come to close grips with this problem (24). His study is based on the final 43 hours of therapy of a neurotic patient. Two judges, using typed transcripts of the interviews, independently rated the therapist on four *a priori* aspects of permissiveness—attentiveness, understanding, gentleness, total communication of acceptance—and on "opening friendliness" at the onset of the hour. Such ratings were available for each hour of therapy, and for smaller units within the hour, "therapist's units" being defined as sequences of the therapist's speech separated from each other by a page or more of the patient's speech. Skin resistance was continuously recorded; the frequency of GSRs exceeding arbitrary slope and amplitude criteria was the dependent variable. To secure a physiological measure within parts of the hour, GSR rate per minute was used.

Overall GSR rate was significantly correlated with whole hour permissiveness (—0.51), initial permissiveness (—0.41), and opening friendliness (—0.52). In hours where the exploration of emotionally significant material was neither very frequent nor very infrequent, and the number of silences was neither very high nor very low, therapist permissiveness was even more highly correlated with GSR rate. "Gentleness" was the single aspect of "permissiveness" which most highly correlated with GSR rate, and "attentiveness" was next. "Understanding" and "overall acceptance" were not related to GSR rate "except as they comprise the factor which may be labelled as gentleness, or a non-punishing manner."

The sensitivity of GSR rate in this study is noteworthy. Whatever the effect of the over-all, more persistent and long-term stability of the patient-therapist interaction, *within-hour* variations in the permissiveness of the therapist had their impact on the patient. If the rated permissiveness of one "therapist unit" deviated by an arbitrary amount from that of the next "therapist unit," the *subsequent* GSR rate was increased. Dittes ingeniously analyzes the data to rule out alternative explanations. The conclusion seems justifiable that within-hour changes in permissiveness had a direct subsequent impact on the somatic arousal of this patient, as measured by skin resistance.

Malmö, Boag, and Smith (47) have contributed an experimental study of the effect on subjects' muscle tension of supportive vs. critical threatening attitudes of the interviewer. Nineteen female psychoneurotic patients were separated into two experimental groups, carefully matched for age and diagnosis. One group was praised for story-telling performance in response to a TAT card, and the other criticized. An elaborate schedule of praise or criticism, and subsequent standardized interviewing was devised. Neck and chin muscle potentials were continuously recorded. Neck muscle potentials yielded insignificant results. Since the chin muscle-lead reflects speech muscle activity, muscle potentials were measured only during silent periods. Various indirect bits of evidence are given to eliminate the objection that the effect of talking was responsible for the significant relationships found between the experimental treatment and chin muscle tension. Significant effects were found, not for muscle tension increases, but for the falling phase of muscle tension after speech. The analyses and findings are difficult to summarize briefly. The general conclusion was that supportive vs. threatening attitudes (praise vs. criticism) did have differential effects

Following praise by the experimenter, the patient's speech-muscle tension fell rapidly. Following criticism, tension tended to remain at high levels. Although cautious and tentative in their arguments, the authors conclude that the lack of significant relationships to the experimental variable of neck-muscle tension is not serious. There is no *general* factor of muscle tension, as the authors correctly point out. They present other evidence, however, suggesting that speech-muscle tension and frontalis muscle tension may be the appropriate skeletal-motor indicators of "arousal level."

The therapist's somatic participation in the therapeutic interchange. It is not surprising to find that the therapist also exhibits a variety of somatic changes related to the course of therapy. The details of the therapist's somatic activity are surprising, however. These details have been supplied primarily by one group of investigators in Boston, who have found a rather startling relationship between physiological change in the therapist and in the patient (5, 12, 22). Apparently, in some cases at least, the therapist responds to the patient's emotional expression as does the patient himself. The implications of this for the therapeutic process need to be worked out.

In 1955, DiMascio, Boyd, Greenblatt and Solomon presented data on the heart rate responses of both therapist and patient (a schizophrenic) for one interview (22). For the first part of the interview, a correlation of $+0.79$ between the patient's and the doctor's heart rate was found; for the latter part of the interview, the correlation was -0.44 . No reason for the shift is given, nor is any interpretation.

In 1956, Coleman, Greenblatt, and Solomon reported an intensive study of one therapist-patient pair through 44

interviews, in which this finding is explored in detail (12). This study has been cited above, and the remarkably sensitive and momentary shifts of the patient's heart rate accompanying momentary shifts among different categories of behavior labelled "anxiety," "depression," "extrapunitive hostility," and "intrapunitive hostility," have been described. The surprising thing is that the therapist's heart rate responded in the same way as the patient's. Like the patient, the therapist's heart rate was highest when the patient exhibited a momentary episode of "anxiety," lowest in momentary episodes of "depression." Like the patient, the therapist's heart rate was intermediate in episodes of "intrapunitive hostility." The only difference between therapist and patient was that the patient's heart rate differentiated "anxiety" and "extrapunitive hostility," while the therapist's heart rate did not. In addition, the heart rate differences among these affective states for the psychiatrist were not as gross as for the patient.

These are surprising data, and may imply that this psychiatrist was responding to the more superficial aspects of the patient's behavior, or it may imply, as the authors seem to feel, a "physiological relationship" between therapist and patient revealing a process of "empathy."

That the "physiological relationship" was not invariant, and that it parallels the "psychological relationship" was shown in additional analyses. It was found that the "physiological relationship" was at a minimum when the therapist was either pre-occupied with his own personal affairs, or when the content of the therapeutic hour was related to the therapist's own unresolved conflicts. Notes dictated by the therapist at the end of each session were reviewed for the presence of statements indicating that the therapist had been disturbed during the hour. The percentage of words de-

voted to statements indicating disturbance was calculated for each hour. This index of disruption of the "psychological relationship" was correlated with a determination of the "physiological relationship" which was based on the extent to which the patient and therapist accorded in yielding faster heart rates during moments of the patient's "extrapunitive hostility" than during moments of the patient's "depression." These two affective states were chosen for study because they enabled comparison in the largest number of interviews. In the six interviews showing the least "physiological relationship," 48% of the therapist's words in his dictated notes indicated disturbance; in the six interviews showing the greatest "physiological relationship," only 24% of the therapist's words indicated disturbance of the "psychological relationship." "In other words," the authors conclude, "evidence of a positive physiological relationship of the therapist to the patient is, in general, found only during those interviews in which the therapist was rarely disturbed by his own preoccupations or by material in the patient's productions or aspects of the patient's transference which would bear on the therapist's unresolved conflicts."

As Coleman, Greenblatt and Solomon imply, these findings raise some important questions. Does the therapist's "empathy" with the patient facilitate his gaining "insight" into the patient? Is the "effectiveness" of the interview affected by the presence or absence of such "empathy?"

In the paper by DiMascio, Boyd, and Greenblatt, already cited (21), additional information is given on the "physiological relationship" between therapist and patient. During the twelve interviews studied, the therapist's electrocardiogram was continuously recorded. As for the patient, the therapist's heart rate was lower in interviews characterized by large amounts of social interaction involving

"tension release," higher in interviews with much "tension," and intermediate in interviews with large amounts of neutral, affectively indifferent, interaction. The therapist's heart rate relationship to these categories, however, was not as pronounced as the patient's. Moreover, the "physiological relationship" held only for "tension." For interactions involving "antagonism," therapist and patient reacted dissimilarly. The therapist's heart rate was positively related ($+.54$) to the relative frequency of "antagonistic" interchange, whereas the patient's heart rate was negatively related ($-.37$). The authors maintain that expression of "antagonism" was "tension-reducing" for this patient.

The already cited experimental study by Malmo, Boag, and Smith (47) contributes further evidence. Like the experimental subjects who were the recipients of praise and criticism, the experimenter exhibited rapidly falling speech-muscle tension after he had administered praise, but his muscle tension did not fall after he had administered criticism. On the other hand, while the subjects' heart rates showed significant variation (in a later stage of the experiment) with the experimental variable, the interviewer's heart rate did not.

ON THE "TRANSACTIONAL" NATURE OF AUTONOMIC RESPONSE

In all the studies reviewed so far, the picture is unmarred. Autonomic and skeletal-motor responses are sensitive indicators, even in the hurly-burly of the interview situation. They are objective. They reveal significant correlations with affect, describe the total effect of a single interview, the variation within interviews, and the long-term course of therapy.

This peaceful picture is partly my fault, because I have deliberately refrained from a few qualifications that emerge here and there in the published reports. I am not certain whether I am

setting up a straw man to defeat when I state that there is a sublime simplicity and faith revealed, in many of the papers, that autonomic and skeletal-motor responses have clear, uncomplicated meaning; a decrease in "physiological tension" means a decrease in "psychological tension," and an increase in the one signifies an increase in the other. It seems as simple as that.

Without meaning to single out any individual or group of individuals, let me cite a few instances where the overt behavior of the individual was at odds with the somatic response, and let us see what the authors make of these apparent incongruities, and how directly and simply interpretations are made.

Anderson finds in his case study (2) that when the content of the patient's speech referred to the therapeutic situation or to the therapist, heart rate tended to be low and stable. Anderson is clear about the meaning: the patient was being defensive, and reducing tension by avoiding self-exposure. No independent data are given to verify this interpretation.

Contrary to Anderson's findings are those of Lasswell (42, 43). Heart rate was found to increase as the number of references to the therapist increased. The proposed reason is that more references to the interviewer signify externalization of the subject's behavior, whereas direct expression of affective feelings to the interviewer would be accompanied by the patient's expectation of retaliation or rejection by the interviewer.⁴ No independent data are given to justify this interpretation.

The patient studied by DiMascio, Boyd, and Greenblatt (21) exhibited a decrease in heart rate, an increase in

heart rate variability, and rise in skin temperature when the patient was expressing hostility and antagonism to the therapist. Little difficulty is found in disposing of this apparent paradox. "The therapist interpreted the patient's expression of antagonism as tension reductive for the patient," and low and variable heart rates accompanied by high and stable skin temperatures is "a physiological pattern which we have repeatedly noted in relaxed subjects." No independent data are given to justify the therapist's interpretation.

At an even greater level of complexity is the case, E. B., reported by Davis and Malmö (16). In "productive" interviews the patient "honestly attempted to explore her emotional feelings and to understand her reactions." During these interviews, she was warm and cordial to the therapist. In the unproductive interviews, the patient was irritable, confused, dependent, evasive, and expressed numerous somatic complaints. She was hostile to the therapist. What happened to her muscle tension? Was it greater as she engaged in effortful and meaningful discussion of personally significant problems, or was it greater as her behavior exhibited disorganization, somatization of affect, and hostility? The former was the case, for her muscle tension was higher in productive interviews than in non-productive interviews. Her hostility, disorganization, and disturbed relationship with the therapist, were not manifested in high muscle tension. The authors propose two explanations. One is that the numerous somatic complaints affected other bodily responses that served as conversion symptoms. The other is that the underlying motivation in the non-productive interviews was to reduce tension and anxiety, and this was reflected in the decreased muscle potentials.

All these interpretations may be correct ones. The point I wish to make is

4. Lasswell also proposes that skin resistance decreases would accompany "unconscious" tension, revealed by slower speech rate. His theoretical reasoning is difficult to follow.

that they reflect a strong faith that a sympathetic-like change inevitably means a parallel sort of increase in anxiety and tension, whereas the opposite changes invariably mean a decrease in anxiety and tension. Nobody seems ever to wonder whether something other than "arousal" or "affect" may be involved, or whether the interpretation of the significance of somatic response involves more than a simple one-to-one formula of relationship. For our present purposes, it should be noted that there is now a note of uncertainty, because the "meaning" of the somatic response has to be interpreted. The somatic response may be related at one time to subtle motivations underlying the patient's behavior to the therapist, but at another time this relation is overshadowed by the relation of the somatic response to the importance of the material being explored by the patient.

In a word, the autonomic response is a part of the total behavior of the subject. The autonomic responses, clear, objective, and sensitive as they seem to be, need to be *interpreted* in terms of the total behavior of the organism. I think most people would agree with this statement when it is made. But this insight, and the definition of the problems invoked by it, are not elaborated except in some psychoanalytic investigations of somatic change during the interview, and in two experimental studies, which will be discussed shortly. The implications are rather serious—even devastating—for those who wish to evaluate the effectiveness of psychotherapy in terms of the course of somatic arousal in a series of interviews, or who wish to "substantiate" the changes "purportedly produced by psychotherapy." Straw man or not, the literature is strangely silent on the "interactional" and "transactional" determinants of autonomic response. So far as laboratory investigations go, the evocation of somatic response upon the

presentation of meaningful, threatening or conflictual symbolic stimuli is an automatic, invariant, and quantitative result. Say, "wife," "mother," or "sex," to a subject, and a response follows almost as invariably as the sun sets at the end of day. So far as many clinical investigations go, the same simplicity of relationship is happily and unquestioningly assumed.

If we observe, as did Dittes in his first paper (23), the skin resistance response to "embarrassing sex statements," we find that the response extinguishes as therapy proceeds. Is this an index of the success of therapy? Is it an index to the successful resolution of sexual conflict? I think there is a strong likelihood that the extinction of the GSR response is *specific to the interactions in the therapeutic situation*, and, in the absence of other confirmatory data, cannot be assumed to generalize to events outside the confines of the consulting room.

In Dittes' remarkably cogent second paper, he places his finger squarely on one extremely important issue (24). He found that a GSR was likely to accompany an "embarrassing sex statement" only during hours when the therapist was rated low in permissiveness and gentleness, either for the whole hour or for the therapist's speech preceding the "embarrassing sex statement." When the therapist was permissive and gentle, GSRs were far less likely to accompany "embarrassing sex statements."

There was only a total of 18 "embarrassing sex statements," occurring in 8 of the 43 hours of therapy. The exploration of sexual material, moreover, occurred only in hours at the upper third or lower third of distribution of the "permissiveness" ratings. A more complete analysis was provided by relating GSR rate to the proportion of time the patient spent exploring any emotionally significant material. Here it was found, for hours of moderate permissiveness

only, that GSR rate was high in hours which had a high proportion of time so spent, and was low when the hours were characterized by low proportions of time so spent. In hours high in permissiveness, however, there was no correlation between GSR rate and the proportion of emotionally significant material! In hours of low permissiveness the correlation was slightly negative. The therapist, if he had wished to do so, apparently could have role-played, and by deliberately varying his permissive attitude could have produced or eliminated this bit of somatic response as an accompaniment of exploration of personally significant material. "The cues," says Dittes, "of discussing emotionally vital matters may evoke GSR only as these cues enter into a kind of *multiplicative* (my italics) relationship with cues of threatening behavior by the therapist." This, to be sure, is one patient, one therapist, and one somatic measure, and the principle needs much confirmation. It is startling to realize how little we know of the specific conditions under which somatic responses can serve as "complex indicators," to use an old phrase.

Some consideration of a recent experiment by Cohen, Silverman, and Burch, may shed some additional light on this problem (11). While this is only a small and preliminary study, an ingenious and important technique is developed for the experimental study of the effect of various therapeutic or non-therapeutic techniques. In this technique, each individual serves as his own control. Some startling results are reported. In the space of a single quasi-therapeutic hour, these investigators were able at will to simultaneously augment GSR reactivity to some symbolic stimuli, and diminish GSR reactivity to others.

The basic technique, in general, was as follows. Each of 5 subjects passively listened to a series of 30 words, in each

of 2 laboratory sessions. On the basis of these two runs, a few words were selected to which the subject had shown high GSR reactivity both times, and other words were selected to which the subject had shown low GSR reactivity both times. An interview followed the second experimental run, at unspecified intervals, and attempts were made to augment GSR reactivity to low GSR-evoking words, and to diminish reactivity to high GSR-evoking words.

In one case, for example, "girl-friend" evoked a consistently high response in the two pre-interview experiments. In the interview, the subject associated to this word, and she was actively supported in her exploration of a homosexual problem centering about another woman living in her apartment. Her feelings, as she explored this area, were reflected; the subject was given a feeling of understanding and acceptance by the interviewer, and so on. By contrast this subject's association to other words ("mother," "cut," "wife," and "hospital") were discouraged after first being encouraged, and the interviewer made no attempt whatsoever to "handle the affect engendered in the course of associations." In the post-interview laboratory session, the GSR reaction to the word "girl-friend" was markedly reduced, whereas the GSR responses to the other four words, where exploration had first been invited and then cut off, were markedly increased.

In a second case, by a similar therapeutic attempt, the GSR response to the word "intercourse" was diminished. The response to the word "vagina," however, was increased, because the authors suppressively instructed the subject not to worry or to think about it. This case certainly shows very little "generalization of extinction," for the words "vagina" and "intercourse" should be rather similar in their impact upon an individual.

All five cases were handled in similar ways, with results all concordant. The results suggest that brief interviews, properly handled, can result in marked changes in somatic reactivity to symbolic stimuli. But the very rapidity of change, and the high specificity of change ("vagina" vs. "intercourse") should make us suspicious. The augmentation or diminution of somatic reactivity did not seem to be a function of the more general interaction between the subject and the interviewer, but were dependent upon highly specific and particular interactions with respect to highly specific word stimuli.

The results of Dittes, and of Cohen, Silverman and Burch, imply that the specific situation is a powerful determinant of the magnitude of somatic response to provocative stimuli. They suggest, I think, that the somatic changes occurring in therapy cannot be utilized as a criterion of the effectiveness of therapy in diminishing somatic effects (or "arousal," or "emotionality") outside of the therapeutic situation. If the disappearance of GSR upon exploration of sexual content is a function of the disappearance of the cues of threat and punishment derived from the therapist, and if GSR can be made to re-appear by the therapist becoming less gentle, more threatening, then we have no guarantee that the patient will not be as somatically aroused as he ever was when he detects "threats" in his environment and interpersonal interactions.

Suppose, then, we resort to "standard" and "universal" stress situations to evaluate the progress of psychotherapy, as was suggested in a pioneering work of Thetford (69). He found that those who had been in therapy recovered more quickly from the frustration engendered by failure at a "mental" task in an independent stress test, than those who had not been

in therapy.⁵ Accepting these results at face value, despite some serious deficiencies in the experimental design and the analysis of results, it should be clear that such a result may be as situation-specific as any other. We cannot use the autonomic response in specific situations as an indicator of the "threat-content" or "arousal value" of that specific stimulus, because we are not yet in a position to measure the "threat-content" of the total situation, and the interaction between patient and examiner. The autonomic response reflects all these aspects; so far we are not able to disentangle the effects. If we attempt to set up situations with "maximal threat-value," to test the limit of adaptation of the organism, we would have to know how the subject perceived the situation, and this would require purely psychological and phenomenological observation. In this sense, autonomic responses cannot be used as a convenient and objective *substitute* for other purely psychological observations.

In another sense, however, which time does not allow us to explore in the detail it deserves, the use of standardized stress situations may be of considerable value, at least in certain psychiatric conditions. In an extensive series of experiments by Malmö and his associates, which Malmö has briefly summarized (45), somatic overarousal to pain or strong auditory stimulation has been shown to differentiate psychiatric patients from controls, and, more importantly, pathological anxiety cases from other psychiatric patients. These data, and some addi-

5. These early results of Thetford's must be viewed with extreme caution. His "control" group was in no sense matched with the "therapy" group on degree of maladjustment. Moreover, the two measures which statistically differentiated the "control" and "therapy" groups at the end of therapy, also significantly differentiated the two groups before therapy was initiated.

tional neurophysiological data, are used to support a theory that defective inhibition, in the neurophysiological sense, underlies clinical pathological anxiety. Insofar as this is true, nonspecific standardized stresses may be used to evaluate the effects of therapy in restoring effective physiological mechanisms of control. These experiments, however, reveal significant group differences only, and neglect (rightfully, in the light of the aims of this series of studies) the strong interactional or transactional effects upon which the present discussion is focused. Indeed, the subtlety of the interactional, transactional, or multiplicative effect is well illustrated by a study from this same laboratory (47), in which 19 psychoneurotic subjects were examined individually on 19 separate days.

Some of these days were "good days" for the examiner, as determined by his diary notes of his feelings and mood; some were "bad days" for the examiner. On the examiner's "good days," the subjects exhibited scarcely any cardiac acceleration as a result of telling a story to a TAT card. On the examiner's "bad days," the heart accelerated significantly, with an average increase of about ten beats per minute. The difference between the "bad day" and "good day" groups was significant at below the 1% level. These effects were not attributable to the following variables: age, diagnostic category, number of words used in telling the story, time examiner took to give instructions, duration of patient's pause before replying to the examiner, deviations from wording of standard instructions, variations in inflection, articulation of words. There was only a hint of what the variable was to which the subjects were responding: the examiner's voice on his very worst days may have been higher in pitch and smoother in texture! The interaction of cues inside and outside the subject seems to be a most subtle matter,

and a very powerful influence in determining somatic activity.

If the therapist or examiner can vary in the "cues" he emits, it is also likely and even obvious that the state of the patient himself can vary. He, too, can have "good" and "bad days." Furer and Hardy, in the paper already cited (30) have a modicum of most interesting data. Their four subjects were stimulated daily with a graded series of pains. After 32 to 49 days, the GSR exhibited adaptation, and could hardly be elicited even to the most intense pain stimulus. The criterion of adaptation was a very rigid one: it was required that there be less than 4% galvanic skin response to the pains of highest intensity for five successive experiments. Even in this highly adapted state, when it seemed that GSR to pain in these experienced subjects was gone forever, it was possible to cause GSR to reappear either by utilizing, or by creating, variations in the general context of the experiment. For one subject, a moderate restoration of GSR response was found when he was preparing material for a scientific meeting. In another subject, the adapted GSR response reappeared when the experimenter deliberately created animosity between herself and the subject. The experimenter who had deliberately created hostility towards herself exhibited an even larger return of response! This same subject showed a return of response when she was questioned about her behavior towards another member of the group.

Even greater complexities of interpretation of somatic response are suggested by psychoanalytic writings. The nature of the evidence is extremely difficult to evaluate, of course, for the voluminous material of extended psychoanalytic sessions is difficult to summarize, and the data upon which summarizing conclusions are based are rarely given, and rarely lend themselves to straightforward evaluation.

Mittelman and Wolff (57) studied continuously recorded finger temperature in 83 interviews. These 83 interviews were part of prolonged psychoanalytic treatment of five patients. The authors claim that in 76 of these interviews they could detect consistent relationships between finger temperature and a large variety of momentarily occurring emotional reactions if six variables were somehow simultaneously evaluated. These they list as (1) the personality makeup of the subject, (2) his attitude towards the analyst, (3) his prevailing mood as expressed in dreams, (4) his changing life situations, (5) the patterning of his aggressive and sexual strivings, and (6) the memories reactivated in analysis. This is a staggering list, and it is a challenge to the experimentalist to translate it into amenable research material. Perhaps most important are the observations that defensive and insulating maneuvers of the patient from the impact of the therapy were quite as successful in preventing the occurrence of vegetative reactions as were "genuine" (their term) states of security, relaxation, and confidence. The authors stress that mere repression of ambivalence and conflict is not sufficient to prevent the occurrence of somatic reactions. The adoption of behavioral defenses and devices, however, which lead to the achievement of emotional security and self-esteem—such as attitudes of superiority, or of being forgiven and cared for—do lead to a reduction of somatic response. The attainment of security through *illusions* of being cared for and of being master of the situation prevented vegetative reactions. If this is true, we would need to be wary, indeed, of accepting diminished somatic response as corroborative of the successful progress of therapy, unless we are willing to accept the "illusion" of mastery and confidence as a successful outcome.

In a similar vein, Reiser, Weiner, and Thaler report (59) in an abstract, that

the development of circulatory changes (heart rate, blood pressure, and ballistocardiograph) in a standardized interview situation, depended on the "object relationship" of the subject and the examiner. Circulatory changes did not appear until a meaningful relationship was established. Even patients with essential hypertension exhibited vascular hypo-reactivity in comparison with healthy young men. "These patients with hypertension demonstrated in the interview 'insulating' defensive techniques, similar to those seen in healthy subjects who were hyporeactors." When, however, a "close relationship" between experimenter and patient was established, physiological hyper-reaction was seen in the interview situation. The authors, with caution, speculate that the "insulating" was "held in check through psychological mechanisms of setting up an 'insulated' object relationship with the examiner."

Other studies and claims could be cited, but perhaps these will suffice. Their import is clear: the "meaning" of the somatic response must be *interpreted* in terms of the transactions of the individual with his environment, and in terms which involve judgments or, at least statements, of "molar" psychological states and behavior. I believe I have pummeled my straw man enough. In the present state of our knowledge, and in the light of our inability to control or evaluate a vast number of relevant psychological, intrapersonal, and interpersonal variables, I do not believe a *substitutive* use of somatic response as an index of change, or of success, in psychotherapy is justified.

I am not, however, as pessimistic about the utilization of somatic responses in the study of the *process* of psychotherapy. provided, again, somatic responses are not used as a substitutive or validating measure. One has only to ponder Dittes' analysis of the "multiplicative" determinants of GSR response or the findings

of Coleman, Greenblatt, and Solomon on physiological consonance between therapist and patient, to see how useful somatic recording in the therapeutic situation can be. For here somatic responses do tell us something, in a fairly unambiguous way, of the process of therapy, providing these phenomena are always integrated by more "molar" and psychological descriptions of the process, which themselves must be substantiated and verified by testing and evaluative procedures appropriate to such "molar" descriptions.

If, however, autonomic and skeletal-motor responses are to be utilized to maximum advantage in the study of the psychotherapeutic process, some attention will have to be given to some rather serious problems, and some theoretical re-organization may be desirable, or at least heuristic.

To adhere to the plan of this conference, I will illustrate some important problems, concepts, and findings primarily with data from the studies performed at Fels in the last decade, with due reference to the work of other investigators where appropriate.

SOME PROBLEMS OF PSYCHOPHYSIOLOGY

The evaluation of autonomic activity.

Three aspects of autonomic activity should be distinguished: (a) tension, (b) lability, and (c) nonspecific activity. The distinction is not merely *a priori* and descriptive. As we shall see, there is some evidence that they have different statistical properties, respond differently to stimulus conditions, and that they may have different relationships to behavior.

Autonomic tension and autonomic lability are simply names for level of function and magnitude of change, respectively. By autonomic tension, then, is meant simply the direct evaluation of the current status of each physiological function in the units of measurement appropriate to that function. The indi-

vidual has a systolic blood pressure of so many millimeters of mercury, a palmar conductance of so many micromhos, a given level of capillary blood-oxygen supply, a given skin temperature, and so on. To compare one function with another, we simply express the measurements in some common unit, such as the deviation from the mean in units of standard deviation, or the related and more convenient T-score. These levels can be studied at rest or as a function of imposed stimuli. When Mowrer *et al.* studied finger-sweating, or Anderson studied a combined index of heart rate and heart rate variability, they studied autonomic tension.

Autonomic lability is a measure of momentary displacement of level as a function of some imposed stimulus. When Mittelman and Wolff reported sharp drops in finger temperature with the introduction of personally significant material, they were studying autonomic lability.

When one desires to establish that individual A is more reactive to a stimulus than individual B, or that a given individual is more reactive to stimulus *a* than to stimulus *b*, one encounters a serious and well-known problem, namely, that the momentary displacement is related to the momentary autonomic tension existing at the moment of stimulation. In general, the higher the pre-stimulus tension (or background excitation), the smaller the magnitude of response, whether the response be computed as a percentage change or as a simple algebraic difference. I have elsewhere considered in some detail the generality of this principle, offered a possible physiological reason for the phenomenon, and proposed an adaptation of regression analysis as a general solution to the problem (39). The method of choice is regression analysis which removes the regression of level attained under "stress" on pre-stimulus level. Removing the regression of

percentage or algebraic change on pre-stimulus level leads to innumerable complications and inaccurate results.

Finally, there is a newcomer to the scene. It was anticipated in many ways earlier in the history of psychophysiology, but it has been formulated as an independently meaningful variable only recently. We called it "spontaneous activity," implying the appearance of autonomic displacement (in heart rate, skin resistance and blood flow) due to internal and "psychologically silent" events (40). By this is meant simply that these autonomic displacements do not occur only as a result of such activities as a shift in perception, or a fleeting change in "affective state." Cohen, Silverman and Burch observed the same phenomenon in skin resistance, and they call it "nonspecific activity" (11).

There are two lines of evidence that this variable deserves to be studied in its own right. In a recent study (40), we showed that the frequency of spontaneous activity, in a subject quietly resting, was related to a variety of aspects of hyperkinetic-impulsive behavior. Autonomic tension, however, was not so related, and, moreover, was but poorly correlated with the frequency of spontaneous activity.

Additional and very significant information is given in the paper by Cohen, Silverman and Burch (11), where still unpublished material is summarized and illustrated. Their "nonspecifics" are small spontaneous skin resistance fluctuations, which have a different relationship to stimulation than do "specific" responses. The authors conceptualize an "arousal continuum" ranging from deep sleep to a moderately alert state and increasing to panic states. When a sedative is administered both nonspecific activity and specific galvanic reactivity to imposed stimuli are found to decrease. When graded doses of epinephrine are administered, the nonspecific activity is found

to continually increase, but specific galvanic reactivity to imposed stimuli first increases and then decreases. Nonspecific activity, in other words, monotonically increases as we move on the "arousal continuum" from the low extreme to the high extreme. Specific galvanic reactivity, however, increases from low to moderate arousal, and then decreases from moderate to high arousal.

Although many additional studies are obviously needed, these two sets of data underscore the heuristic value of assuming the relatively independent status of the variable of "spontaneous" or "nonspecific" autonomic activity. Dittes' GSR measure, it should be noted, was a measure heavily weighted with nonspecific activity.

The spontaneous activity seen at rest is not merely chaotic autonomic change. Individual differences in the amount of spontaneous activity are quite reproducible upon 48-hour re-test (40). Spontaneous activity increases during some forms of "stressful" activity, and recently Dr. N. T. Welford, Mrs. Lacey, and I have found that it may decrease systematically in simple motor tasks, where extreme regularity of tracking motions are required.

Figures 1 and 2 illustrate the phenomenon of nonspecific spontaneous cardiac oscillation in two subjects at rest. I present them to illustrate the fact that there is a sort of complex periodicity present, that "spontaneous activity" is unmistakably different from stimulus-evoked activity, and because later I want to discuss an experiment which will require understanding of the nature of cardiac oscillation.

The figures are derived as follows. It is a commonplace that continuous records of heart rate exhibit variation from one beat to the next. It is relatively rare to find two successive R-R intervals that have precisely the same value. Sometimes this continuous shift is related to

respiration (respiratory or sinus arrhythmia). The heart accelerates, beat-by-beat, as the subject inhales, and decelerates, beat-by-beat, as the subject exhales. If one carefully observes this repetitive wave-like form, one also observes that the peak heart rate (minimum R-R interval) achieved in each inspiratory cycle itself exhibits oscillation, increasing and decreasing almost periodically, and with a relatively long period. Similarly, the trough (minimal) heart rates, found at the end of expiration, increase and decrease, with a long periodicity. So far

as we have been able to observe, this variation in what we call the envelope of peaks, or the envelope of troughs, is not related to any detectable changes in respiration. In subjects not showing marked respiratory arrhythmia, the low frequency oscillation is clearly discernible.

In Figures 1 and 2 we have plotted against time each point of inflection in the peaks and the envelope of troughs, for two subjects sitting quietly at rest for 16 minutes. They are extreme cases. Figure 1 shows an individual with marked oscillation of high and variable

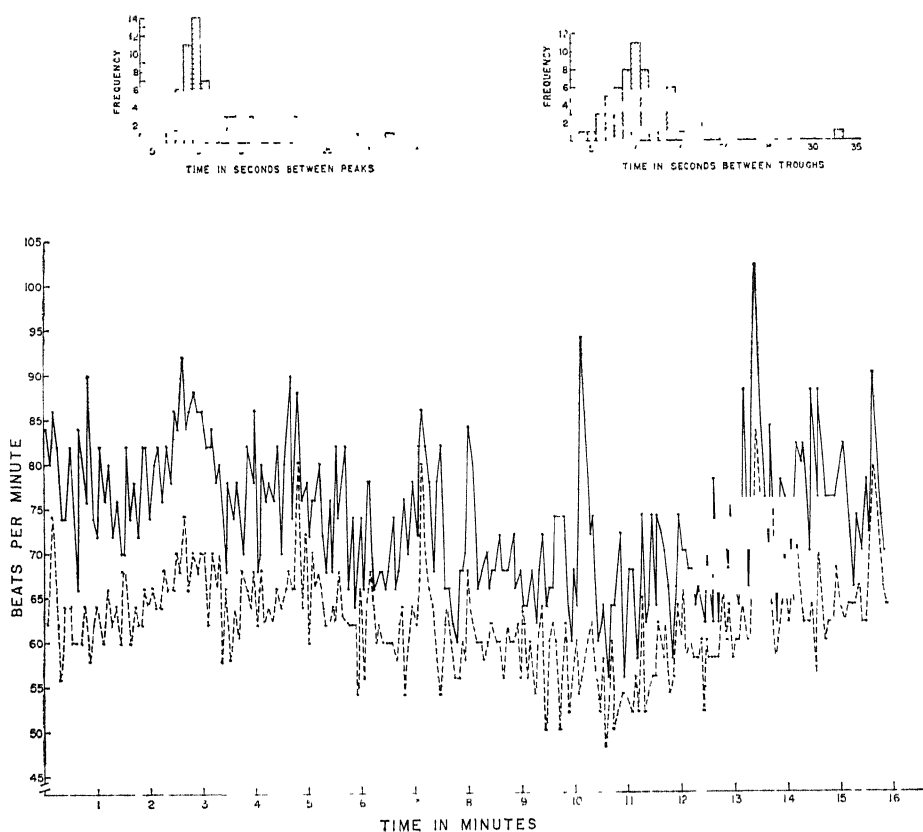


FIGURE 1

Showing the variation of heart rates in the "envelope of peaks" (solid line) and the "envelope of troughs" (broken line) in a subject who is quietly resting in a comfortable armchair. Each point of inflection of a beat-by-beat record of heart rate is shown. In the inserts are shown frequency distributions of the intervals of time between successive points of inflection. This is a "labile" subject, showing large amounts of spontaneous activity.

amplitude, with a modal period between points of inflection of about 10 seconds. There is discernible a fairly regular second-order oscillation. Such individuals, we found, are hyperkinetic-impulsive. Figure 2 shows an individual with much less marked amplitude of variation, a longer modal period (for peaks) and a wider spread of period. Such individuals we found to be low in hyperkinesis-impulsivity. Our measures of hyperkinesis were primarily: (a) reaction time, (b) maintenance of a set to respond, so that reaction times did not abruptly increase as the foreperiod was increased, and (c) the frequent occurrence of responses to signals in the periphery of the visual field to which the subjects had been instructed *not* to respond. In a later section, we will return to these findings, and the theory underlying them, and their possible applicability to the therapeutic situation

On the intercorrelations among autonomic measures. One of the most crucial

issues in psychophysiology concerns the surprisingly low intercorrelation among measures. In our work with a variety of noxious stimuli and a simple variety of autonomic variables, we have consistently found matrices of inter-correlations in which the majority of correlations approached zero. In our most recent study, for example, (41), where adult women were subjected to a variety of situations which successfully activated the autonomic nervous system, palmar conductance tension scores had no significant correlation with systolic blood pressure, heart rate, and variability of heart rate; palmar conductance lability scores were also independent of all other lability scores. So far as these data went, palmar conductance, heart rate, and variability of heart rate were essentially independent measures, whether autonomic tension or autonomic lability scores were used. This is essentially what we have found each time we looked into the matter. Similarly, in our study of spontaneous activity

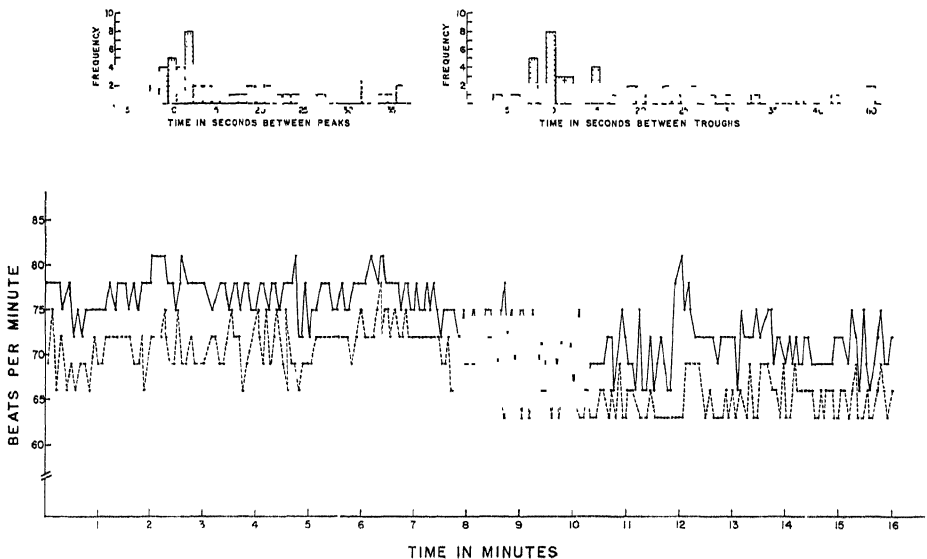


FIGURE 2

Drawn as was Figure 1, but this is a "stable" subject with long periods of time between points of inflection, and low amplitude of variation of the envelopes of peaks and troughs.

(40), we found that the frequency with which cardiac variation exceeded an arbitrary magnitude was poorly correlated with the frequency of similarly defined GSR variation. For a group of adult women we secured a barely significant (at the .05 level) correlation of .36, which dropped to .32 on re-test. For a group of college males, the correlation was +.08.

This is not a unique experience. Wenger (72) reports the intercorrelations for a very wide range of variables for an N of 488 aviation cadets. It is difficult to find a correlation above .30. With a 20 X 20 matrix of intercorrelations, most of the variables of which are resting autonomic tension scores, there are only 10 correlations higher than $\pm .20$. The highest is .44, and that is between two intimately related variables, salivary output and salivary pH. In a similar 20 X 20 matrix for 62 children, there were only 38 correlations higher than $\pm .20$ (71). Other extensive modern studies of relationships show that no single physiological measure correlates well with others, and that no single measure can serve as an index to the state of other measures or to the total "arousal" of the organism (3, 9, 10, 17, 19, 48, 62, 68).

There is, of course, probably no such thing as *the* correlation between any two autonomic variables. Correlations probably change as a function of many variables, not yet understood. Even in the best of circumstances, however, where substantial intercorrelation is seen, for example, upon the injection of epinephrine (9), the correlation is low enough so that one measure cannot be easily substituted for other measures.

Complete absence of response in one variable. A simple and convincing example of the difficulty of using a single arbitrarily chosen variable as an indicator of "arousal" or "affect" is found where

one variable completely fails to show any response, while another simultaneously recorded variable shows considerable impact of the stimulating condition. This sort of occurrence, in my experience, is not so rare as many seem to think.

For the purposes of this conference, I went through some old data, and made a simple tally of such discrepancies, for two commonly used variables, palmar conductance and heart rate.

One set of data consisted of the records of 94 boys and girls, aged 6 to 18 years, who for each of six years have come to our laboratory for a longitudinal study of physiological response to a very painful stimulus, the cold pressor test. This requires immersion of the foot in ice-cold water (4°C) for 75 seconds. The children hate this test, and, with the younger ones particularly, it is easy to see the desperate conflict between the desire to get the foot out of the water, and the desire to satisfy the socially defined requirement to suffer in silence. After relaxation to steady physiological levels, the subject is quietly informed that he (or she) will be required to immerse his (or her) foot in the cold water after one minute. There follows one minute of what we think may justifiably be called apprehension or anxiety. We label it the "alerting period." After one minute, the cold pressor test is administered. Following the cold pressor test, the subject is allowed to recover to resting physiological levels, and the entire procedure is then repeated. The details of this experiment are given elsewhere (36, 41).

For the 94 records, gathered in the course of one winter, 10.6% of the cases exhibited cardiac acceleration during the first alerting period, but completely failed to show palmar conductance response. For the second alerting period, 11.2% of 89 cases (5 individuals could not endure the stress, broke down, and cried) showed the same failure of palmar con-

ductance to reveal response, while their heart rate was simultaneously increasing.

Conversely, in other cases, palmar conductance did show an impact of the alerting period, while cardiac rate failed as an "indicator." In the first alerting period, 14.9% of the individuals exhibited this pattern; in the second alerting period, 10.1% of the group.

In approximately 21% to 25% of the cases, then, palmar conductance and heart rate yielded completely discordant answers to the very simple question: was this individual affected by the stimulus-condition?

During the cold pressor test itself, similar discordance was revealed, with smaller frequency. For the first of the two tests, palmar conductance failed to respond, while heart rate was simultaneously showing marked and typically sustained acceleration, in 2.2% of the cases. For the second cold pressor episode, this pattern was found in 5.7% of the cases. Similarly, cardiac rate failed to indicate "arousal," while palmar conductance was simultaneously showing response, in 7.8% and 6.8% in the first and second cold pressor tests respectively. In the cold pressor test itself, then, the incidence of such complete discordance ranged from 10% to 12.5%. There were no age or sex differences in the results.

The result is not limited to this age span, nor is it dependent on the fact that most of the group had had repeated yearly experience with the test. This was established by reviewing the records of a group of 68 adult women (the mothers of these children) for the first time they were subjected to a cold pressor test. They received only one such test, the repeat determination upon recovery from the first test being omitted. In the alerting period, there were 7 cases (10.3%) where palmar conductance did not respond, but heart rate did. There were, in addition, 4 cases (5.9%) where palmar conductance responded, but heart

rate did not. Complete discordance was found, then, in 16.2% of the cases during the alerting period. During the cold pressor test itself, palmar conductance failed in 10 cases (14.7%), who were simultaneously exhibiting cardiac acceleration, and heart rate failed in an additional 10 cases who were simultaneously exhibiting skin resistance changes. Thus, complete discordance was found in 29.4% of the cases during the cold pressor test itself.

In the literature on the psychophysiology of the interview this situation has been noted several times. Mowrer *et al.* (58) note that even in their small sample of cases studied by the finger-print stain technique of measuring palmar perspiration, there were two cases where palmar perspiration was a completely inadequate measure. One case, despite numerous somatic complaints, exhibited absolutely minimal perspiration. In the chemical laboratory where he was employed "he could handle polished surfaces without leaving finger marks on them, surfaces which most people would not dare to touch."

Similarly, in the paper by Mittelman and Wolff (55), there was a group of patients who, despite obvious behavioral evidence of intense disturbance, exhibited only very minor changes in finger temperature in the interview situation.

In this last cited paper, one of the sources of this sort of discrepancy among physiological indicants is obvious. Most patients in this study were already exhibiting psychosomatic disorders clearly related to finger temperature. They all had pre-existing complaints of cold hands under stress, and some had frank Raynaud's disease. The group of patients who showed only minor temperature changes in the stress of interview did not have such psychosomatic complaints. A great deal of the data showing striking concordance between finger temperature change and the emotional content of the

interview is due to the selection of a physiological indicant related to already existing somatic disorder.

This is in accord with the principle of "symptom specificity" as developed by Malmö and his collaborators (49, 50, 51), which states "... that in psychiatric patients presenting a somatic complaint, the particular physiological mechanism of that complaint is specifically susceptible to activation by stressful experience," a principle partly derived from the data already cited on muscle potential changes during interview. These authors, too, are careful to make clear in several places, in the studies cited in an earlier section on indicant functions, two things: (a) they selected patients with symptoms "which could be studied appropriately by means of electromyography," and (b) some patients may clearly reflect clinical progress in muscle tension measurements, others may not.

In these findings from the laboratory and the interview we have a strong hint of individual differences in the organization of somatic responses: for some individuals, one or more physiological measures are minimally reactive in a wide variety of life-situations (Mowrer *et al.*), and other measures are maximally reactive (Mittelman & Wolff; Malmö *et al.*) in a wide variety of life-situations. One thing seems clear: if there is a specific somatic complaint, it would be the part of wisdom to choose to measure somatic responses related to the physiological mechanism underlying that complaint. A second thing seems also clear: wholesale reliance on measurement of a single variable may lead us astray. It might be concluded, at first glance, that the problem is not too serious; that, for example, if one chooses skin resistance as an index of bodily arousal, one may simply lose a few cases. The true situation, however, is much more complex. Because autonomic variables have low intercorrelations, the different physio-

logical responses of a given individual to a given set of stimulus-conditions can be plotted as an uneven profile of reaction: High reactivity or arousal may be exhibited in one or more physiological variables, low reactivity in other variables, and average reactivity in still others. These hierarchies or patterns of response are reproducible, and vary lawfully as a function of the individual, and as a function of certain variations in the stimulus-conditions.

At present, we have only a very inadequate understanding of the principles underlying such patterning of somatic responses. In attempting to summarize what is known in this area, I will use the terminology and the categories developed in the most recent paper on this subject from our laboratory (41). The relevant facts can be arranged in three categories: (a) intra-stressor stereotypy of response; (b) inter-stressor stereotypy; and (c) situational stereotypy. There is a fourth category—symptom specificity—which has been mentioned briefly earlier in this section, but this category is omitted in what follows.

Intra-stressor stereotypy. By intra-stressor stereotypy is meant that in response to a given stressor episode, an individual exhibits a reproducible pattern of response. In other words, the low intercorrelations commonly found among autonomic measures are due in part to idiosyncratic organization of the pattern of response. This is one of the basic problems encountered when physiological measures are used as indicators of "arousal."

In our first study of this problem (35), a small group of women was subjected to a simple stressor, namely, reciting as quickly as possible as many words as the subject could remember that begin with a stated letter of the alphabet. Under the circumstances of the experiment, the subjects became quite anxious and em-

barrassed because of their "obvious" inability to do well at this task for even as long as one minute. This stressor was imposed approximately once a month over a nine-month period. Using a ranked percentage measure of physiological response, it was found that each subject responded with an individually characteristic pattern of response, with statistically significant reproducibility over the nine-month period.

In a second study (36), the immediate test-retest reliability of the pattern of response was evaluated in a different population of subjects and with a different stressor. A large group of boys and girls between the ages of 6 and 18 was studied. The stressor used was the very painful cold pressor test. Striking patterning of response was again found. Whether a subject was categorized as average in reactivity or "arousal," or as a hyper-reactor or hypo-reactor depended entirely on the physiological response that was used for the assay of the individual. One physiological variable—heart rate, for example—could place a given subject two standard deviations above the mean in reactivity, whereas another simultaneously recorded variable—skin resistance, for example—could place this same individual one standard deviation below the mean. Statistically, the pattern of response was again found to be reproducible.

In a third, unpublished, longitudinal study, it was found that the pattern of response to the cold pressor test is reproducible over as long a period as four years, again in a group of children between the ages of 6 and 18. In this study autonomic lability scores were used, as defined earlier in this paper. Age and sex factors were statistically eliminated where necessary.

It will be useful to discuss one of the analyses of this study in some detail, to give some concrete idea of one of the statistical bases for the conclusion that

idiosyncratic organization of response patterns does exist, and to give some preliminary data on the nature of the distribution of the tendency to exhibit intra-stressor stereotypy.

The coefficient of concordance was used to evaluate the reproducibility of the pattern of response. This is a sort of average rank-order correlation, ranging from 0 to +1.00, expressing the agreement of pattern obtained over the 4 years. The frequency distribution of the individually computed coefficients of concordance was determined, and compared with the frequency distribution expected on the basis of chance. The details of this procedure, and appropriate references, are given in an earlier publication (41). The obtained and chance frequency distributions are shown in Figure 3. They are significantly different at below the 1% level of confidence, by the Kolmogorov-Smirnov test for comparing an obtained with an hypothetical distribution (64).

Over a period of four years, then, even in the face of all the physiological and psychological changes seen in maturing children, there is a strong tendency for individuals to reproduce an idiosyncratic pattern of somatic response to the cold pressor test. For some, this tendency is quite strong. As is shown in Figure 3, 27% of the group showed coefficients of concordance greater than 0.70. For others, the tendency is modest, and for still others the pattern of response varies randomly from one occasion to another. The data suggest quantitative individual differences in the tendency to exhibit stereotypy of response patterns. Some individuals seem so constituted that they rather rigidly exhibit the same hierarchy of response from one occasion of measurement to another; others respond randomly, showing now one pattern of response, now another.

It should be noted in Figure 3 that there is a uniform shift to higher levels

of concordance. At low levels of concordance there is a deficiency of observed frequencies, compared to expected or chance frequencies; and, at higher levels of concordance, there is an excess of observed frequencies. If only a proportion of subjects truly had a tendency to reproduce a pattern of response, we would not expect a deficiency in absolute numbers of cases at lower levels. Repeated observation of this uniform shift to higher concordance levels is, to date, the main evidence supporting the notion of a quantitative continuum of individual differences in the tendency to stereotypy of response patterns.

While this hypothesis is in accordance with all the evidence to date, there is an alternative hypothesis. Perhaps all individuals have a characteristic and modal response pattern, one that would appear more frequently than any other in an infinitely large series of trials; and, per-

haps, the relative frequency of occurrence of this modal response pattern is approximately the same for all individuals. Using a finite series of trials, however, we are in effect sampling each individual a limited number of times, and, as a statistical result of sampling, a frequency distribution is generated in which individuals appear to be distributed along a continuum ranging from randomness to fairly rigid stereotypy. A decision between the alternative hypotheses can be readily made experimentally, by determining whether the classification of an individual as "stereotyped" or "random" is reliable from one set of trials to another. This experiment is now underway, with each "set of trials" numbering 12, and with re-test intervals of 1 week, 1 year, and, hopefully, 4 years. The "trials," however, are not repetitions of one stressor, but of a series of different stressors, because of additional find-

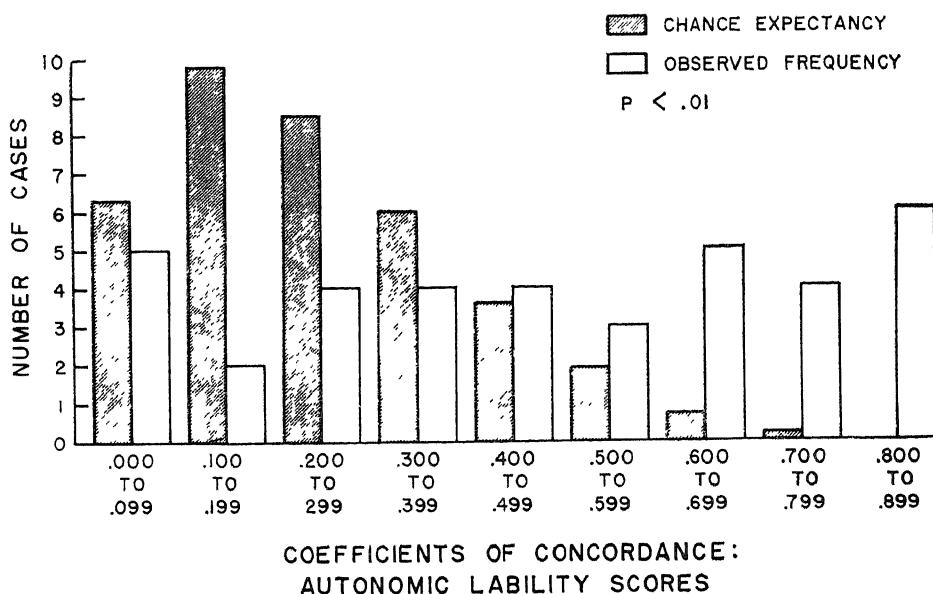


FIGURE 3

Showing obtained and hypothetical frequency distributions of coefficients of concordance, computed to show the reproducibility of response-patterns to the cold pressor test over a period of 4 years in a group of 37 children between the ages of 6 to 18 years. Age and sex effects on autonomic lability scores were removed.

ings on idiosyncratic organization of response patterns, to which we now turn.

Inter-stressor stereotypy. The significance of these findings of reliable differential somatic response patterns is greatly extended by evidence showing that the pattern obtained using one stressor tends to be reproduced in other stressor episodes with radically different physiological and psychological demands. This phenomenon we previously called autonomic response specificity. In order to provide a reasonably systematic nomenclature, however, the term "inter-stressor stereotypy" is now used.

In our first study of this problem (38), 85 male college students were subjected successively to 4 stressor episodes, while skin resistance, heart rate, and beat-to-beat instability of heart rate were measured. The four stressor episodes were hyperventilation, the cold pressor test, mental arithmetic, and the letter association (or word fluency) test described earlier. The results clearly supported the hypothesis of inter-stressor stereotypy.

In a second study (41), adult women were used as subjects, and systolic, diastolic, and pulse pressures were added to the physiological battery. The list of stressor episodes was modified slightly: apprehension of painful stimulation, mental arithmetic, word fluency, and the cold pressor test were used. The results of the first study were verified and extended. Some of the data and results of this verification study are presented in the following paragraphs, because they provide a fairly inclusive summary of the main facts and principles obtained to date.

In Figure 4 are shown four representative profiles of reaction, using autonomic tension scores. It is clear that there is striking fractionation of response on each of the five occasions of measurement, for each of the four subjects, and that the idiosyncratic patterns of response are

closely reproduced from one occasion of measurement to another. For case No. 32, for example, palmar conductance is very high, but heart rate instability is very low. For case No. 31, heart rate levels indicate very high levels of bodily arousal, whereas systolic blood pressure levels indicate low bodily arousal. Any pattern of response can be illustrated by one or more of the 42 cases in this study. It should be noted here that the results are not explicable by hemodynamic principles. That is, an obtained pattern of high blood pressure and low heart rate is not to be explained as an example of reflex cardiac slowing in consequence of momentary hypertension, (a) because, in the circumstances of this study, cardiac levels and blood pressure levels were positively, although poorly, correlated, and (b) because the same pattern is obtained whether blood pressure is high (say, during the cold pressor test), or low (say, at rest). Similar arguments hold for obtained patterns of heart rate and heart rate variability. Moreover, the same sort of reasoning holds for autonomic lability scores. In the original report of this study (41), detailed intercorrelational matrices are given, which may be consulted for verification of these statements.

That these four cases are not selected and unrepresentative ones is shown in Figure 5, in which are shown the expected and obtained frequency distributions of individually computed coefficients of concordance. There is a marked and obvious discrepancy. The two frequency distributions differ significantly at far below the 1% level of confidence. Moreover, a noteworthy total of 39 out of 42 individuals showed statistically significant coefficients of concordance.

Using autonomic lability scores, similar results are found, at lower levels of confidence. This contrast between autonomic tension scores (physiological levels attained) and autonomic lability scores (physiological changes) is one we have

repeatedly observed. We have no explanation for it as yet. One real possibility is that there is more unreliability in autonomic lability scores because of the inevitable processes of statistical estimation (of population means, population standard deviations, and population correlations) that enter into the determination of measures of change of physiological functions.

In Figure 6 are shown four representative cases, illustrating, again, striking fractionation of response and reproducibility of the pattern of response over four stressor episodes, using autonomic lability scores. Case No. 6 should be noted particularly. This subject shows marked

variation both in the degree of activation from one stressor episode to another, and in the magnitude of fractionation of response. In both these respects, this subject is rather atypical. Nevertheless, note that the rank order of the physiological responses tends to be maintained. Even in the cold pressor test, for example, in which all the physiological measures show hyper-reactivity, heart rate variability response is greater than heart rate response, which is, in turn, greater than palmar conductance response; this is the same state of affairs characterizing two of the other three stressor episodes. The result is a significant coefficient of concordance (0.70).

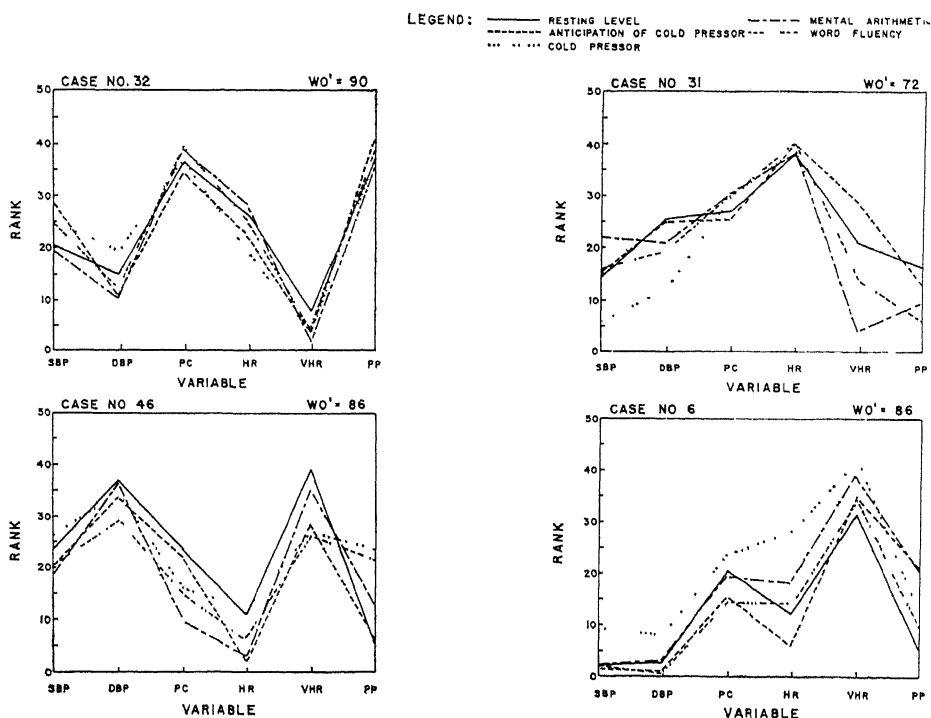


FIGURE 4

Four examples of idiosyncratic response patterns (autonomic tension scores) reproduced over five occasions of measurement. Physiological variables are on the abscissae: SBP, DBP, and PP are systolic, diastolic, and pulse pressure; HR and VHR are heart rate and beat-to-beat variability of heart rate; PC is palmar conductance. The ordinates are ranks, showing the relative position of the subject in the total group of 42 subjects. Wo' is the coefficient of concordance, corrected for continuity.

In Figure 7, obtained and chance frequency distributions are compared. The familiar deficiency of cases at low levels of concordance and excess of cases at high levels is again seen. The difference between the two distributions is significant at the 3% level. Nine of the 42 individuals (approximately 21%) yielded statistically significant coefficients, three of which were below the 1% level of confidence.

Thus, in five studies of autonomic activation by physiological and psychological stressors, and in the study of spontaneous resting autonomic activity, we have uniformly found the two basic phenomena of fractionated or patterned autonomic response, and tendency to reproducible profiles of reaction. In the data presented in an earlier section concerning heart rate and palmar conduc-

tance responses, we found that patterning of response can even extend to complete failure of response in one physiological variable, while another simultaneously recorded variable indicates "arousal." The results seem to have considerable generalizability, for various subject populations have been used—maturing children of both sexes, male college freshmen, and adult women—in experiments in which details of experimental design, recording techniques, and mathematical modes of evaluation of magnitude of response have all differed. For the kinds of stressors we have employed, and for the physiological measures used, we believe that the evidence is conclusive that individual differences in the organization of response hierarchies exist, and that these differences are reproducible over long periods of time and from one stres-

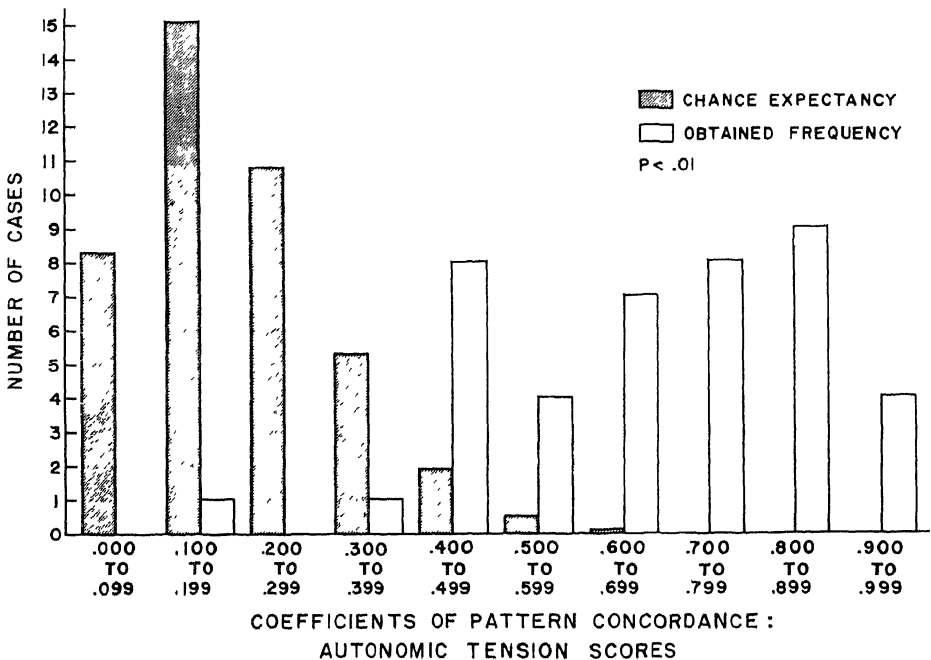


FIGURE 5

Showing obtained and hypothetical frequency distributions of coefficients of concordance which show the reproducibility of response patterns over 5 occasions of measurement, in a group of 42 adult women. Age effects on autonomic tension scores were removed.

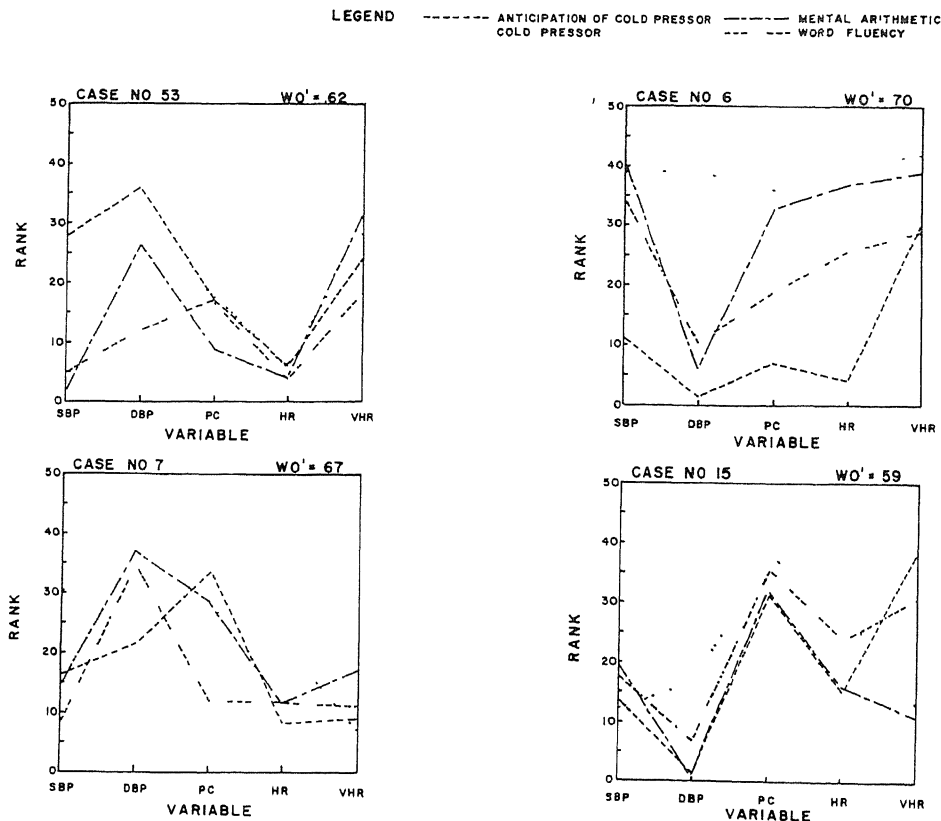


FIGURE 6

Four examples of idiosyncratic response patterns (autonomic lability scores) reproduced through 4 stressor episodes. Abscissae and ordinates are labelled as are those in Figure 4.

sor to other different stressors. We believe also that the evidence is suggestive of quantitative individual differences in the strength of the tendency to reproduce response patterns, either over time, or from one stressor to another.

Before going on to a discussion of situational determinants of response patterns, it seems desirable to consider the implications of these findings for the conduct of experiments in which the "emotional responsiveness" or the "affectivity" of an individual has to be assayed, or in which the "arousal values" of different situations or stimuli are to be evaluated. We have approached this problem in two different ways, by study-

ing what we have called "indicant concordance," and by using peak or maximal reactivity as an index of "arousal."

Indicant concordance. The facts of inter-stressor and intra-stressor stereotypy of response patterns show that each physiological variable yields a different "arousal value." No single variable, therefore, can assign a unique "arousal value" to a stimulus, nor can it assign a unique "responsiveness" to an individual. A single physiological variable, however, still might be useful in quantitatively differentiating among stimulus-situations, if the various physiological measures rose and fell together in the concordant manner shown by case No. 6 in Figure 6.

The results for this individual show that with a few small inversions, all physiological measures agreed in ordering the four stressors from least arousing to most arousing, the order being: anticipation of painful stimulation, the word fluency test, mental arithmetic, and the cold pressor test. As already discussed, however, this subject is atypical in several respects, and is a member of what is probably a very small subset of individuals, as is made clear in the original report of this experiment (41) where the statistical mechanics underlying the emergence of response stereotypy are considered. Only rarely do different physiological measures agree so well in ranking the "arousal value" of different stimulus-conditions.

In a quantitative study of this problem, we computed coefficients of con-

cordance for the 42 adult women in the study, this time in such a way that the resulting coefficient expressed the agreement among the physiological measures in rank-ordering the stressors from least "arousing" to most "arousing." A coefficient of concordance computed in this way we called a "coefficient of indicant concordance." For autonomic lability scores there was no indication of significant indicant concordance. The ordering of the stressors in terms of "arousal value" for the individual depended entirely on the physiological variable selected. For autonomic tension scores, the answer was only a bit more favorable. Figure 8 shows the results, comparing obtained with chance frequency distributions of the coefficients of indicant concordance. While the obtained frequency distribution is signifi-

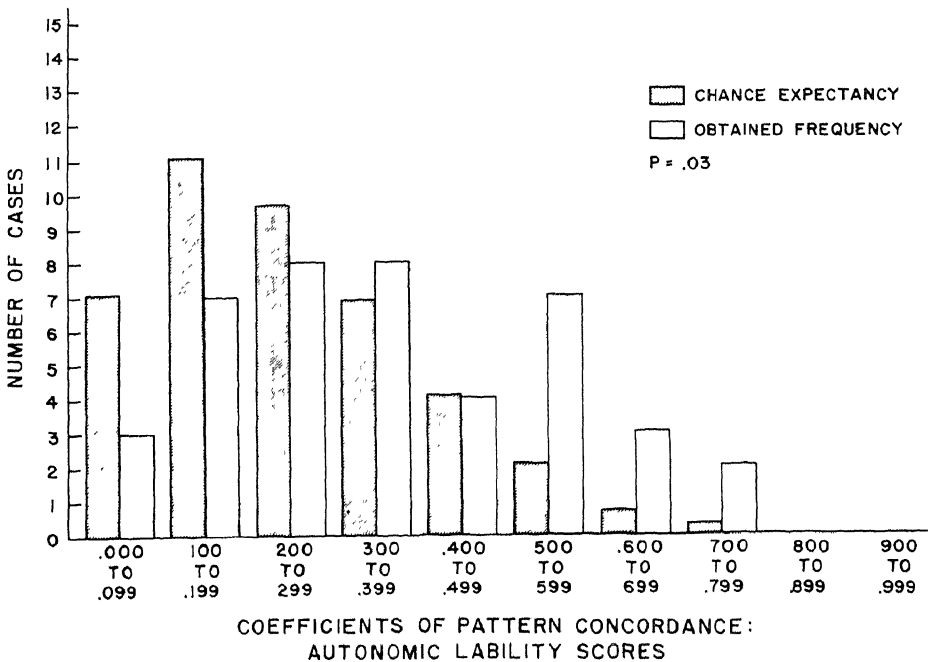


FIGURE 7

Showing obtained and hypothetical frequency distributions of coefficients of concordance which show the reproducibility of response patterns over four stressor episodes, in a group of 42 adult women. There were no age effects on autonomic lability scores. (See reference 41).

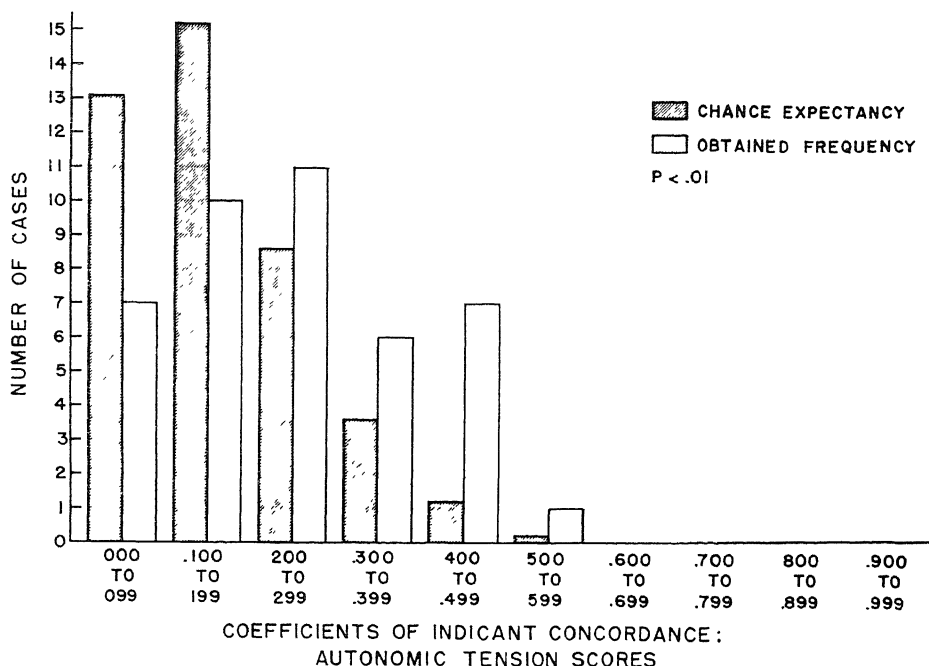


FIGURE 8

Showing obtained and hypothetical frequency distributions of coefficients of indicant concordance, which show the agreement among 5 physiological variables in rank-ordering the "arousal value" of 5 stimulus-conditions (see Figure 5), in a group of 42 adult women. Age effects on autonomic tension scores were removed.

cantly different from the chance distribution, even the highest coefficients are disappointingly low.

Within the limits of this experiment, then, there is no evidence for indicant concordance of high enough magnitude to justify reliance on single measures to sort out differential "arousal values" of different stimuli. The notion of over-all autonomic arousal, however, is so fixed in our psychophysiological traditions that it may be useful to ask if this conclusion is in agreement with other data.

While the question has not really been studied in any detail, some clarification results if *group* indicant concordance and *individual* indicant concordance are distinguished. By group indicant concordance is meant that on the average, for the group as a whole, each of a variety of physiological measures may show

similar effects, or bear similar relations to stimulus-parameters, for example. Group indicant concordance has been shown many times, and is one reason, I think, for the wide-spread practice of arbitrarily selecting a single physiological variable as an indicator function. The demonstration of group indicant concordance, however, obviously does not imply that for the individual the same effects will be shown by any one variable as by any other. Few published experiments lend themselves to a suitable analysis of the difference between the two forms of indicant concordance. There are two experiments, however, which clearly bear on the issue.

Hovland and Riesen (34), as mentioned earlier, showed that both galvanic skin response and vasoconstriction were linearly related to graded intensities of

painful electric shock and of acoustic loudness, thus clearly demonstrating group indicant concordance. They noted, however, that in individual cases, the two physiological variables were frequently discordant. In studying this phenomenon, they computed individual product-moment correlations between the two physiological responses, for each of eight subjects. These correlations were: .32, .35, .44, .54, .56, .68, .68, and .71, all with a positive sign. These correlations were based on sizeable N 's, ranging from 31 to 61. Using appropriate Z transformations, we tested the significance of the differences among the 28 possible pairings of these 8 correlations. Seven pairs were significantly different at below the 5% level of confidence; of these, 2 were significantly different at below the 1% level. These results, then, are in accord with our finding of systematic individual differences in the organization of response patterns, and agree in indicating a striking lack of individual indicant concordance, although, it should be emphasized, the same data show group indicant concordance.

In a somewhat different approach, Malmo and Davis (48) showed that several physiological measures—blood pressure, heart rate, muscle potentials, and respiration—all showed steady increases in time ("rising gradients") during the performance of mirror tracing. The slopes of these gradients were each positively related to the adequacy of the motor performance. Malmo and Davis interpret the results to mean that the slope of the rising gradient reflected the arousal value or the motivational value of the situation. Accepting this interpretation, this experiment, too, clearly demonstrates group indicant concordance. When the intercorrelations among the different physiological measures were computed, however, only moderate values were found, most correlations falling below +.50. Again, then, fractionation

of response is shown, and it is clear that different physiological variables yield discordant results for the individual.

For the individual, then, (the proper study of psychotherapy) it appears that reliance cannot be placed on single physiological variables. None of these studies have utilized stimuli or criterion measures directly relevant to the verbal symbolic nature of psychotherapy. It is a matter of some importance, then, to inquire into individual indicant concordance in the interview situation itself. How high is it? Do some individuals show concordance, and others not? These questions should be answered soon, before the psychophysiological study of psychotherapy progresses much further. The experimental evidence so far obtained, however, constitutes grounds for serious pessimism that a satisfactory solution to the problem of assessing individual "arousal" (or the "arousal-value" of individual stimuli) will be achieved until methods are devised to take into account each individual's reaction profile. We have made one feeble attempt to accomplish this, by using maximal reaction of the individual as an index of the individual's "affectivity."

The use of peak or maximal arousal. Given the facts of reaction-profiles, it is tempting to average reactivity in a number of autonomic "channels" as a measure of autonomic "arousal." Although such a technique has been used successfully by Sines (65), the practice seems dubious. Consider an individual for whom we have measured three autonomic variables. Suppose the three reactivity measures, in units of sigma deviation from the mean, are +1.5, 0, and -1.5. His average would be zero, meaning that he is of average reactivity. This conclusion would be a rather absurd one.

More complicated schemes—for example, using multiple regression analysis

and assigning beta weights to the various measures—are available, of course, where there is a specific criterion measure and when large samples are used. Such techniques, however, are not easily adaptable to the individual case, and would require large numbers of repeated observations on each individual.

In an earlier report (37) we attempted to evaluate general or overall autonomic reactivity by using the maximum response of an individual as the index of "arousability," "responsiveness," or "affectivity," no matter in what variable or in what stressor this maximum response was found. Thus, two subjects were called equally reactive if the reaction profile of one subject showed a maximum T-score of 70 (2 standard deviations above the mean) in palmar conductance response to the cold pressor test, and the other subject's reaction profile also showed a maximum score of 70, but in heart rate response to hyperventilation. Obviously, this technique is flatly empirical, and lacks a sound physiological basis. Empirically, however, these subjects could be considered to be equally reactive within the limits of the experiment. We showed that the physiological measures, taken one at a time, did not yield consistent or significant correlations with the Rorschach form-color index of "emotional-ity." When we correlated maximum autonomic reactivity with this index, however, we secured a correlation of $+0.47$, significant at the 2% level of confidence.

In an unpublished study with Dr. Vaughn Crandall, however, using a different population of subjects, a different Rorschach examiner, and in a different social context (with respect to the interaction between Rorschach examiner and examinee) we were unable to verify these results. We were about to forsake the technique, but two recent reports from other laboratories support its use.

Vogel, Baker, and Lazarus (70) found that of several techniques of evaluating autonomic reactivity that they used, the maximal arousal technique was the only satisfactory one, enabling correlation of physiological reactivity with a variety of psychological variables. Judging from an incomplete report (66), Spence also has found the technique to be useful. After several fruitless attempts to use GSR as an index of drive in simple conditioning experiments, he was successful in establishing a correlation of peak autonomic lability (he measured skin resistance and heart rate responses to a mild noxious stimulus) with conditioning acquisition curves.

These reports give some reassurance that the idea has some worth, but a great deal of work needs to be done to arrive at a sound understanding of the significance and usefulness of the technique, and of the circumstances under which it can be used. In particular, it should be noted that one of several complicated and hidden assumptions in the technique is that each of the physiological measures used correlates with the criterion measure in the same sense, with the same algebraic sign. That this may not be true under all conditions will be shown in the next section, which deals with the third, and final, category under which the principles of response patterning are discussed in this paper.

Situational stereotypy. The search for differential patterns of bodily response in differently named affective states was abandoned quite early in the history of psychophysiology, with results that were generally conceded to be disappointing. This experimental enterprise has been effectively renewed in recent years, with rather dramatic results, that are, moreover, of interest and importance to the problems involved in the psychophysiological study of the therapeutic interview. In brief, changes in *average* or *modal*

response patterns have been produced by changes in the stimulus used, or have been related to qualitative differences in the affective experiences aroused by appropriate stimulating conditions. One very important line of investigation finds significant differences among states of "anger-directed-inwards," "anger-directed-outwards," "fear" and "anxiety," theoretically linked to a biochemical variable—the ratio of epinephrine to nor-epinephrine secretion (3, 29, 62).

The most dramatic examples come from R. C. Davis' laboratory (17, 18, 19, 20), where changes in modal response patterns to a variety of simple stimuli have been established. In these studies—in which subjects are stimulated by warmth or cold, look at affectively toned pictures, tap telegraph keys, or listen to auditory stimuli—differential response patterns are secured in which there are not only changes of quantitative emphasis among the physiological measures, but in which startling and unexpected (in the light of our traditions) reversals of *direction* of physiological change occur.

For example, in a response pattern which includes marked vasoconstriction and pronounced palmar sweating, tradition would lead us to expect cardiac acceleration. Instead, Davis' P-pattern shows vasoconstriction, increased palmar conductance, and cardiac deceleration and is the modal pattern obtained when college students look at affectively toned pictures.

Findings such as these constitute a most serious embarrassment to simple activation or mobilization theories of autonomic nervous system function. In my opinion they are likely to lead to extremely fruitful insights, the realization of which has been blocked by tradition, although, as we shall see, there are several experiments in the early literature that should have led to a determined search for such response patterns long

ago, and should have been taken as early clues to the significance of these patterns for concomitant or subsequent behavior.

The relationships of situational stereotypy to intra-stressor and inter-stressor stereotypy are not at all clear yet. This topic has been briefly discussed elsewhere (17, 41). In addition, it should be noted that the investigations of inter-stressor and intra-stressor stereotypy have not utilized experimental situations which produce striking qualitative or "directional fractionation" of response (as we may call the phenomenon) such as found in Davis' P-pattern, for example.

In the study mentioned earlier as being in progress, in which the reliability of classification of individuals as "random" or "stereotyped" responders is being studied, the 12 stimulus-situations employed were chosen to represent as wide a range of stimulus-situations as seemed expedient. Quite by accident, we have stumbled on some extremely suggestive examples of "directional fractionation" of response. In one pre-test of the large-scale experiment, 15 subjects were exposed to the finally selected series of 12 stimulus-situations. It was noted that several of the situations had particularly striking differential effects on the direction of cardiac change. Some situations produced cardiac deceleration almost uniformly, whereas simultaneously recorded blood flow and skin resistance were changing in the expected "sympathetic-like" direction.

In Figure 9, some of these results are shown, for four of the stimulus-conditions, arbitrarily named "visual attention," "empathic listening," "thinking" and "withstanding pain." In the "visual attention" task, the subject was required to silently note the colors and patterns produced by flickering photic stimulation, with instructions that he would be asked to describe them later. None of the subjects was nauseated by the stimulus, and none responded with any marked degree

of affect. In the "empathic listening" task, the subject listened to a 60-second dramatic rendition, tape-recorded by an experienced and effective professional actor, of the thoughts and feelings of a badly injured and dying man. The subject was instructed merely to listen attentively and empathically. The "thinking" test was a mental arithmetic test, requiring the rapid solution of a series of simple addition and multiplication problems. The problem was given orally, and the subject was required to solve the problem "mentally." As soon as he announced an answer, another problem was immediately given, and the process repeated as many times as could be accomplished in two minutes. The "withstanding pain" test was the cold pressor test.

The main point to note in Figure 9 is the "directional fractionation" of the response pattern. While all four stimulus-

conditions produced increases in palmar conductance, the direction of heart rate changes differed in the four conditions. Thus, in the "visual attention" task, 13 of the 15 subjects exhibited cardiac deceleration, whereas, in the cold pressor test, 14 exhibited cardiac acceleration. By Cochran's Q test (64), the differences among experimental situations in the obtained response patterns were significant at below the .001 level of confidence. When all 12 stimulus-situations were used for this statistical test, the confidence level was .01. We are not, therefore, capitalizing on chance in Figure 9: only four experimental conditions are represented there for reasons of convenience and of clarity of presentation.

It probably would be wise to stop at this point, indicating merely that we corroborate Davis' finding that response patterns may vary as a function of the stimulus-conditions, and that these re-

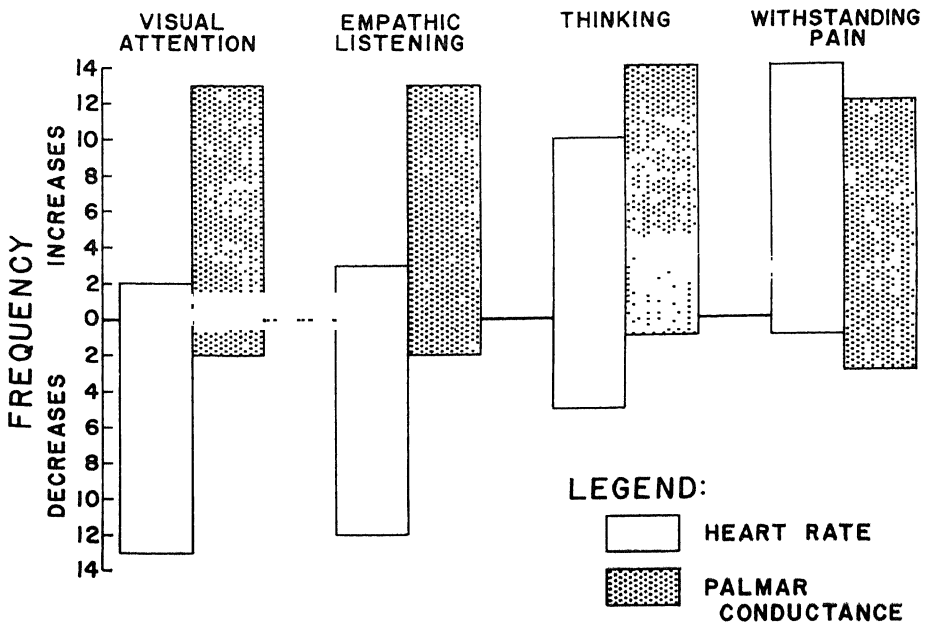


FIGURE 9

Showing the different effects of 4 stimulus conditions on 15 subjects in evoking increases or decreases in palmar conductance and heart rate.

sponse patterns fit no known or accepted concept of autonomic functioning. Certainly these patterns deny the usefulness of a concept of sympathetic *versus* parasympathetic function, and, just as certainly, argue against an easy use of "sympathetic-like" changes as indicators of "arousal" or of "behavioral intensity."

It is tempting, however, to speculate a bit about such results, especially if hypotheses can be derived from the speculations that can be put to experimental test, the results of which might be of use to the psychophysiological study of the therapeutic interview. The speculations that follow derive from the notion that it might be fruitful to consider whether these patterns of response have any utility to the organism.

Response patterns and organism-environment relations. Perhaps it should be emphatically repeated that much of what follows is speculation, based, to be sure, on some factual data, but lacking the firm base of a series of integrated experimental tests.

In what may well turn out to be the single most pregnant sentence in the current literature on the psychophysiology of the autonomic nervous system, Davis *et al.* say (19): "Something of the original stimulus differential is preserved up to the end of the chain." While the data in Figure 9 agree with the nature of the findings upon which this statement was based, we do not agree that autonomic responses are "the end of the chain." Moreover, we are inclined to speculate that it is not the "stimulus differential" that is being preserved in a response pattern—which would make the long-latency autonomic responses into something like an information-transmitting system—but that the dealings ("transactions") of the organism with the environment are being facilitated or hindered.

In a previous publication (40) we have included a review of some of the

neurophysiological and anatomical evidence that argues against considering autonomic responses as "the end of the chain." Only a small part of the evidence, and the attendant argument, will be repeated here. It is truly an historical accident, and a matter merely of pedagogical definition, that the autonomic nervous system is considered to be purely an effector system. The earliest investigations of the autonomic nervous system noted that reflex chains were completed by visceral afferent fibers, leading from the activated viscus back to the central neuraxis. These visceral afferent fibers, however, did not present the structural and functional peculiarities of the efferent limb of the autonomic nervous system, and were arbitrarily excluded from textbook discussions of the autonomic nervous system, precisely because they were no different in function from other sensory or afferent nerves. In current neurophysiology, the homologies between the classical sensory pathways and the visceral afferent pathways are again being stressed. In psychology, visceral afferent pathways have been largely neglected, and play a role only in the James-Lange theory of emotion, which emphasized that when some autonomic effectors became active, the organism could sense and perceive their activity. However, in addition to those visceral afferents that give rise to perception, there are those that subserve reflex functions only, not capable of being perceived.

Recent neurophysiological evidence shows that at least one visceral reflex pathway may be of great significance for the control of behavior. The pressure-sensitive receptors in the carotid sinus, long known to participate in cardiovascular homeostasis, have now been shown to exercise tonic *inhibitory* control over cortical electrical activity and over spinal motoneurone activity. The carotid sinus is strategically located to serve as an

immediately responsive device for the detection of changes in heart rate and blood pressure; and, in the performance of this task, it is joined by pressure-sensitive receptors in the aortic arch, and probably also by other pressure-receptors in the walls of the heart itself. Earlier work established the extreme sensitivity of these receptors to even very small changes of intra-sinusal pressure.

An increase in heart rate or blood pressure, then, is very likely to lead to *inhibitory* effects on cortical activity, and on motor activity. As is emphasized by Callaway and Dembo (7), moreover, such "sympathetic-like" changes may lead to inhibitory effects on sensory and perceptual events. Far from being the "end of the chain," then, these *autonomic responses become stimuli* to internal receptors, whose activation may well lead reflexly to changes in the relationship of the organism to the environment, in terms both of the organism's accessibility to environmental inputs and the organism's motor outputs.

Now, *a priori* at least, the four experimental conditions of Figure 9 can be ordered along a continuum. At one end, where cardiac deceleration is the rule, the subject is required primarily to note and detect the environment: he is asked to note colors and patterns, to listen attentively. At the other end, where cardiac acceleration is the rule, one can at least speculate that the opposite of environmental intake is called for. The subject listens to a very brief announcement of an arithmetic problem, and then tries to concentrate on internal symbolic manipulations aimed at solving the problem; or, he experiences rather severe pain and cold, in which physiological mechanisms to decrease environmental intake would be very useful. In a sense, then, the acceleration or deceleration of the heart could be considered to be something like an instrumental act of the

organism, leading either to increased ease of "environmental intake," or to a form of "rejection of the environment."

This is a very fancy interpretation indeed! There is, however, some corroborative evidence in the existing literature, and I have an experiment from our laboratory to report, which tends to verify the interpretation. These bits of evidence are reviewed in the following paragraphs.

Darrow, almost 30 years ago, provided an early clue that unfortunately was never pursued. In a very detailed survey of literature (13), and in an independent experiment (14), he reported differences in the response patterns produced by "sensory" and "ideational" stimuli. Simple sensory stimuli, calling for "no extensive association of ideas" (simple environmental intake?) resulted in heart rate *decelerations*; on the other hand, either noxious stimulation or a sequence of activity requiring "associative processes" (rejection of the environment?) produced heart rate *acceleration*. Moreover, "sensory stimuli" were more effective than "ideational stimuli" in exciting peripheral changes such as skin resistance and vasoconstrictive responses.

It may also be recalled that the early Wundtian formula (61) equated pleasant stimuli (which the organism wants to "take in"?) with cardiac deceleration, and unpleasant stimuli (which the organism wants to reject?) with cardiac acceleration. Canestrini, cited in (8), found this formula to hold even in the neonate.

It is certainly relevant that Davis' P-pattern—great sweat gland activity, peripheral vasoconstriction, and cardiac deceleration—"is produced in its purest form in male subjects presented with a picture of a nude female" (17). I am being only slightly facetious when I suggest that this is a clear-cut example of an individual wanting to "take in the environment."

These observations essentially only show a relationship between the direction

of heart rate change (which seems to vary independently of other aspects of the sympathetic-like "arousal mechanism") and dimensions of the stimulus situation. By inference, they show a relationship of the direction of heart rate change to the motivation of the organism in dealing with his environment. Pointed studies are badly needed to substitute direct experimental verification for this inference.

In the meantime, another aspect of the problem is demonstrated in three other experiments. In these studies, it is not the relationship of stimulus conditions to autonomic response patterns that is studied, but rather clues are provided to some of the physiological and behavioral consequences of over-all or fractionated autonomic activity.

In the already-cited paper by Callaway and Dembo (7), the hypothesis is advanced that various methods of producing over-all sympathetic-like activity, underlying which is an hypothetical common neurophysiological factor labelled "central sympathomimetic activity," result, by unspecified physiological mechanisms, in decreased effectiveness of environmental events. For example, in this paper, and in other papers which are therein cited, they found that size constancy deteriorated, and that electromyographic and skin resistance responses to sensory inputs decreased when massive sympathetic-like effects were experimentally produced. While their hypothesis and findings are similar to or congruent with ours, there are two important differences. First, they feel that the diminution of environmental influence is not total. They speak rather of "narrowed attention" in which only the more peripheral stimuli (peripheral in space, time, and relevance of the stimuli to the task at hand) are rendered less effective. Secondly, they are careful to state that the correlation between "narrowed attention" and increased "central sympathomimetic

activity" does not imply a direct causal relationship. We are emphasizing, not an hypothetical "central sympathomimetic activity" correlated with "narrowed attention" via some more basic mechanism, such as stimulation of the brain-stem reticular formation, but a part mechanism directly in series, so to speak, with the chain of physiological events eventuating in behavior—namely, increase or decrease of carotid sinus inhibition of cortical electrical activity, spinal motoneurone activity, and possibly sensory afferent activity. We believe, moreover, that the data of the experiments by Callaway and his collaborators suggest over-all diminution of environmental influence, rather than the "narrowed attention" they conceptualize.

Darrow *et al.* (15), moreover, have provided some evidence that the fractions of the autonomic response pattern have different physiological correlates (or effects). They showed that two "sympathetic-like" changes had different effects on electroencephalographic changes produced by sensory stimulation. Palmar conductance increases were "excitatory" and were positively correlated with alpha blocking. Blood pressure changes, on the other hand, were "homeostatic" and were negatively correlated with alpha blocking; *i.e.* they resulted in inhibitory effects on cortical excitation.

Finally, Mrs. Lacey, Dr. N. T. Wellford, Dr. William D. Thompson, Jr., and I have just completed an experiment in which we attempted to demonstrate that cardiac *deceleration* and skin conductance *increase* had similar effects on behavior. This experiment was a follow-up of a previous demonstration (40) that resting rates of spontaneous cardiac and sudomotor activity were each independently correlated with the height of the stimulus-generalization gradient as determined in the situation described by Brown and his collaborators (6). In this situation, the subject is instructed to raise

his hand from a telegraph key as quickly as possible when, following a preparatory signal, a white light in the center of the visual field is momentarily flashed. When any one of six other lights (placed peripherally to the right and left of the center light at 8°, 16°, and 24° of visual angle) is flashed, the subject is instructed not to react. Under the conditions of the experiment, in accordance with the principles of stimulus-generalization, subjects do respond occasionally to the peripheral lights. We interpreted these unwanted "false" responses as instances of impulsive motor action, which, although contra-indicated by instructions, the subject could not inhibit (40).

In the present experiment, we desired to demonstrate, first, that the probability of evoking a "false" response was a function of the momentary changes of heart rate depicted in Figures 1 and 2; and, second, that the probability was also determined by skin conductance activity, but in an opposite direction. Specifically, our hypothesis was that when heart rate was momentarily high, the probability of a false or impulsive response to a peripheral light would be less than when the heart rate was momentarily low, and, on the other hand, that when skin conductance activity was momentarily high, the probability of a false response would be higher than when skin conductance activity was momentarily low. Increase in cardiac rate, then, was hypothesized to be inhibitory, increase in sudomotor activity, excitatory, following the reasoning already set forth, and the physiological demonstration by Darrow *et al.* of a similar phenomenon in alpha block, already discussed.

The technique was as follows. An apparatus was so designed that the peripheral lights would flash only on pre-determined heart rates. This was accomplished by having the voltage output of the cardiometer (which is proportional to momentary heart rate) close

a relay in series with the stimulus lights, if it reached or exceeded a pre-set voltage value. After observation of the maximal and minimal heart rates shown by the individual, circuit constants were set (by appropriate variable resistors) so that the lights would be activated only if the momentary heart rate equalled or exceeded some high value, or equalled or was lower than some low value. The high and low values, it should be clear, were determined individually for each subject, and were meant to approximate the peaks and nadirs of the oscillation of heart rate, depicted in Figures 1 and 2. Only two peripherally located lights were used, 8° to the right and left of the centrally placed light. Each of these lights was activated 12 times, 6 times on cardiac "highs," 6 times on cardiac "lows." The subject, therefore, had 24 opportunities to execute, or to refrain from executing an "impulsive" response. The sequence of right and left lights, on high and low heart rates, varied in pre-arranged haphazard order. Between each pair of trials employing peripheral lights there was a minimum of three trials employing the central light. Plantar conductance (right sole to left sole) was continuously recorded, simultaneously with heart rate. Preparatory signals were not used, because it was impossible, of course, to predict exactly when a cardiac "high" or a cardiac "low" would appear. The foreperiod interval, therefore, could not be controlled. Instead, the subject worked for 20 minute periods, during which he was to maintain a set to respond or not to respond, depending on the stimulus. Five minute rest periods were interspersed between work periods.

The inability to predict when a cardiac "high" or a cardiac "low" would occur led to a serious complication. In early versions of this experiment, it proved impossible to have the apparatus simply "command" that, for example, the left hand light be activated when the next

cardiac "low" appeared, because a very lengthy interval of time, sometimes extending for as long as two minutes, would ensue before the cardiac "low" occurred. Subjects soon become drowsy and bored with such lengthy intervals between stimuli. Therefore, the apparatus was modified so that it worked essentially as follows. A programming device, actuated by pre-punched tape, gave the "command" to flash a peripheral light on, say, the heart rate corresponding to a cardiac "high," if the cardiac "high" occurred within a pre-determined interval of time—3, 4, or 5 seconds. If the appropriate momentary heart rate did not occur in this time-interval, the "command" was automatically "aborted," and the central light, to which the subjects were instructed always to respond, was flashed, following which the same cycle of events was automatically repeated until the desired cardiac "high" or cardiac "low" occurred. In this way the subject was kept continually busy and alert. The programming tape provided a prearranged haphazard order of inter-trial intervals, ranging from 6 seconds to 15 seconds, with an average value of 12 seconds.

Our first aim, to demonstrate that the probability of a "false" response was different on cardiac "highs" and "lows" was almost, but not quite, blocked by the necessity for this "command-abort" system. The net effect of this system was that the number of trials in which the central white light was flashed could not be controlled. For example, following a trial employing a peripheral light on a cardiac "high," we might have wanted to intersperse four trials with the central light before trying for a peripheral light on, say, a cardiac "low." But it might require five additional trials before the cardiac "low" occurred. In consequence, different subjects received different numbers of center-light trials, and, within each subject, the numbers of center light

trials preceding cardiac "lows" and "highs" were not identical.

Now, it is no surprise that the number of center-light trials preceding a peripheral-light trial is a potent factor in changing the probability of a "false" response to a peripheral light, because successive center-light trials lead to a set to respond, or a sort of behavioral inertia. It turned out that this factor was much more important than we had anticipated. By a simple inference, however, the data show, we think, that a false response to a peripheral light is less likely to occur if a light is flashed on a cardiac "high" than if it is flashed on a cardiac "low."

The facts were as follows. The average number of false responses (for 30 subjects) was approximately 5.5 when lights were flashed on cardiac "highs"; on cardiac "lows," the average number of false responses was also 5.5! This diminishing equality, however, clearly is explicable by the intervening number of center-light trials, as follows.

(1) For all peripheral-light trials in which false responses occurred, the average number of preceding center-light trials was approximately 9. For those peripheral-light trials in which false responses did not occur, the average number of preceding center-trials was approximately 7. This small difference of 2 was statistically significant, P being below .01 by the Wilcoxon paired replicates test (64). (2) When the same analysis was done for peripheral lights on cardiac "lows" only, similar results were obtained. The average number of center-light trials preceding a trial in which a

6. In very early versions of the experiment, using inadequate technique, we thought we detected the opposite effect, with more false responses occurring at high heart rates. (See the Discussion appended to 40). This conclusion was in error, and, of course, is contrary to theory. The source of the error was primarily the number of center-light trials.

false response was evoked was approximately 9; the number preceding a peripheral-light trial in which a false response was not evoked, was approximately 6.75, and the P value was less than .01. (3) For trials on cardiac "highs" only, the corresponding figures were 9.5 and 7.5, with P less than .02. These three analyses show that a surprisingly small increase in the number of center-light trials preceding a peripheral-light trial significantly increased the probability that a "false" response would be evoked by the presentation of a peripheral light. For convenience in later discussion, this effect will be called the response-set effect. Now, (4) *the average number of center-light trials preceding peripheral-light trials on cardiac "highs" was significantly greater than the average number of center-light trials preceding peripheral-light trials on cardiac "lows,"* the figures being 8.8 and 6.7 respectively, with P less than .01 by the paired replicates test. This effect, of course, was unplanned. It came about, we think, in the following way. The nature of the cardiac oscillation is such (see Figure 1 in particular) that, at intervals, quite high heart rates occur, that are considerably above the average. Corresponding drops to excessively low heart rates are not as likely to occur. Cardiac nadirs, in other words, do not vary as radically as cardiac peaks. In setting maximal and minimal heart rates into the "command" portion of the apparatus, we unwittingly chose, in general, truly maximal rates that occur less frequently.

It should now be obvious that the response-set effect should have resulted in a greater number of responses on cardiac "high" trials. This effect was not seen, however, because the average number of false responses was the same on cardiac "highs" and "lows." The obtained number of "false" responses on cardiac "highs," in other words, is less than expected, in the light of the effect

of response-set. Why? We think that one fair inference is that the attainment of cardiac "highs" produced momentary motor, or sensorimotor, inhibition, in line with the theory already outlined.

An alternative explanation would be that the momentarily occurring cardiac "highs" accompanied or produced more "alertness," more "arousal," so that the subject did not "mistake" a peripheral light for the central light. By this simple "arousal" theory, however, increased GSR activity should also result in diminished incidence of false responses. The data, however, show the opposite effect: GSR activity *increased* the probability of a false response. This was established in the following way.

Because the response-set effect is so powerful, we considered separately peripheral light trials which had been preceded by more than the median (for each individual separately) number of center-light trials, and those that had been preceded by less. Since plantar conductance had been continuously recorded during the experiment, as well as heart rate, it was a simple matter to determine whether there was or was not skin resistance activity (arbitrarily designated as a GSR exceeding 1000 ohms). Since a GSR takes time to occur, and occupies relatively extended intervals, compared with momentary heart rates, we counted all GSRs that began within an arbitrarily selected 10 seconds preceding the light-stimulus.

For presentations of the peripheral lights on cardiac "highs," with below the median number of preceding center-light trials, there was a significant positive association between the incidence of a "false" response and the occurrence of concomitant or just-preceding skin resistance activity (chi-square of 4.6, significant at between the .05 and .02 levels, with 1 degree of freedom). For light-stimuli presented on cardiac "lows," (with less than the median number of

preceding center-light trials) there was no such effect. Using peripheral-light trials for which the number of preceding center-light trials was greater than the median, there were no significant associations of GSR activity and false responses, on either cardiac "lows" or cardiac "highs."

The data, then, for trials with below the median number of preceding center-light trials, indicate that the two "sympathetic-like" changes—increases in heart rate and skin conductance—have correlations with "false" responses that are opposite in sign. Like Darrow and his collaborators, we find that skin conductance increase is excitatory, and cardiac acceleration is inhibitory. On cardiac "highs," when there is momentary inhibition, the excitatory effects of GSR activity can summate with the inhibitory effects to decrease the inhibition. On cardiac "lows," when inhibition is already removed, GSR activity had no significant effect. Since these findings held only when peripheral-light trials with below-median number of preceding center-light trials were studied, the response-set effect was a powerful interactive influence, capable of hiding other effects.

Because these analyses are quite complicated, and the inferences from them are proportionally subject to error, the interactive and summational effects of skin conductance and cardiac activity were analyzed in a very different manner, using all trials, and including the response-set effect. It was noted that subjects apparently varied in the importance of the response-set effect in determining their responses. For each subject, a frequency distribution was made of the number of center-light trials preceding peripheral trials. For some subjects, false responses were made primarily only after many center-light trials; for other subjects, false responses were more evenly distributed, and could occur as easily after 3 preceding center-light trials as

after 10. We computed, for each subject, a "response-set ratio," which was the proportion of false responses that occurred when the preceding number of center-light trials exceeded the median number for that subject. A ratio of .60, then, means that 60% of the individual's false responses occurred after a greater-than-median number of center-light trials. Each subject had two response-set ratios, one for cardiac "high" trials, and one for cardiac "low" trials. Having determined these response-set ratios, we categorized each subject as a GSR "stable" or a GSR "labile," depending on whether spontaneous resting GSR activity was, respectively, below or above the median for the group, using a technique previously described (40).⁷

For peripheral light presentations on cardiac "highs," the median response-set ratio of 15 GSR "stables" was 0.61; for 15 GSR "labiles," it was 0.51. The difference was significant at below the .05 level, by Wilcoxon's unpaired replicates test (64). GSR "stables," then, needed more assistance, so to speak, from the response-set effect, before they made a false response on cardiac "highs," than did GSR "labiles." For trials on cardiac "lows," however, the response-set ratios were 0.58 for both "stables" and "labiles."

Thus, this very different form of analysis shows the same results as the preceding analyses, with skin conductance activity being excitatory and overcoming the presumably inhibitory effect of a high heart rate, but having no detectable effect when heart rates are low, this latter state of affairs itself being excitatory.

In summary, we can say that the results support (but do not constitute clear experimental proof of) the concept, and the underlying neurophysiological theory,

7. These data were secured from a 15-minute rest period that preceded the experiment.

that skin conductance increase is excitatory, whereas increase of cardiac rate is inhibitory of this simple transaction of the organism with the environment. The pattern of response obtained when recording skin resistance and heart rate may reveal occasions when the individual is "open to his environment" and ready to react to it, or, conversely, when the individual is not "open," and, indeed, instrumentally "rejects" the environment.

In this sort of investigation, the autonomic nervous system is seen, not simply as a source of indicant functions, but as part and parcel of the organic determination of behavior.

Even more subtle effects are demonstrable. Callaway and Dembo showed (7) that "central sympathomimetic states" decreased the adequacy of learning in a probability-learning experiment. The gross sympathetic excitation produced by administration of methamphetamine led to "narrowed attention" and an inability to use the cues provided by relative frequency of events in the environment.

Some applications to the study of psychotherapy seem obvious, especially if psychotherapy is viewed as a learning process, and as one characterized by social interactions, some of which can be arranged on a continuum from "environmental intake" to "environmental rejection with accompanying internal elaboration."

Whether this concept is right or wrong, or somewhere in between, it is clear that our traditional and traditionally-derived concepts—such as homeostasis, energy mobilization or arousal, sympathetic vs. parasympathetic function, and over-all sympathetic activation—fail to encompass much data. We believe, also, that it would be most valuable, in continuing the enterprise of studying psychotherapy by psychophysiological means, to search for new concepts—such as visceral afferent feedback—to guide investigations

into the means by which the autonomic nervous system influences, modulates, and directs behavior. In addition, one might hazard the guess that considerable progress may be made in the psychophysiological evaluation of psychotherapeutic process and outcome if "transactional" influences on autonomic activity—in the sense defined in the text of this paper—were to be given at least as much emphasis as has been given to indicant functions of the autonomic nervous system. Finally, in using autonomic measures as indicator functions, we believe that particular attention should be given to the following principles: (1) Clear distinctions should be made among three dimensions of autonomic activity, namely, tension, lability, and spontaneous activity; (2) Appropriate and defensible mathematical techniques must be used in arriving at measures of lability. Simple difference or percentage measures of change are rarely, if ever, defensible. This topic has been touched upon only lightly in this paper. It is discussed in detail elsewhere (39). (3) One measure of somatic "arousal" cannot serve as an index to the state of other measures. Even at best, the intercorrelations among autonomic measures is low. (4) So far as present evidence goes, *individual* indicant concordance is so low that single autonomic measures cannot be used to unequivocally rank-order the "arousal-value" of different stimuli for a given individual, or the "arousability" of different individuals.

REFERENCES

1. Alexander, F. Psychoanalytic study of a case of essential hypertension. *Psychosom. Med.*, 1939, 1, 139-152.
2. Anderson, R. P. Physiological and verbal behavior during client-centered counseling. *J. counsel. Psychol.*, 1956, 3, 174-184.
3. Ax, A. F. The physiological differentiation between fear and anger in humans. *Psychosom. Med.*, 1953, 15, 433-442.

4. Bixenstine, V. E. A case study of the use of palmar sweating as a measure of psychological tension. *J. abnorm. soc. Psychol.*, 1955, 50, 138-143.
5. Boyd, R. W., & Di Mascio, A. Social behavior and autonomic physiology: A sociophysiological study. *J. nerv. ment. Dis.*, 1954, 120, 207-212.
6. Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, 41, 52-61.
7. Callaway, III, E., & Dembo, D. Narrowed attention. A psychological phenomenon that accompanies a certain physiological change. *Arch. Neurol. & Psychiat.*, 1958, 79, 74-90.
8. Carmichael, L. (Ed.) *Manual of child psychology*. New York: Wiley, 1954.
9. Clemens, T. L. Autonomic nervous system responses related to the Funkenstein test. I. To epinephrine. *Psychosom. Med.*, 1957, 19, 267-274.
10. Clemens, T. L. Autonomic nervous system responses related to the Funkenstein test. II. To Mecholyl. *Psychosom. Med.*, 1957, 19, 363-369.
11. Cohen, S. I., Silverman, A. J., & Burch, N. R. A technique for the assessment of affect change. *J. nerv. ment. Dis.*, 1956, 124, 352-360.
12. Coleman, R., Greenblatt, M., & Solomon, H. C. Physiological evidence of rapport during psychotherapeutic interviews. *Dis. nerv. System*, 1956, 17, 2-8.
13. Darrow, C. W. Differences in the physiological reactions to sensory and ideational stimuli. *Psychol. Bull.*, 1929, 26, 185-201.
14. Darrow, C. W. Electrical and circulatory responses to brief sensory and ideational stimuli. *J. exp. Psychol.*, 1929, 12, 267-300.
15. Darrow, C. W., Jost, H., Solomon, A. P., & Mergener, J. C. Autonomic indications of excitatory and homeostatic effects on the electroencephalogram. *J. Psychol.*, 1942, 14, 115-130.
16. Davis, F. H., & Malmö, R. B. Electromyographic recordings during interview. *Amer. J. Psychiat.*, 1951, 107, 908-916.
17. Davis, R. C. Response patterns. *Trans. N. Y. Acad. Sci.*, 1957, 19 (Series II), 731-739.
18. Davis, R. C., & Buchwald, A. M. An exploration of somatic response patterns. Stimulus and sex differences. *J. comp. physiol. Psychol.*, 1957, 50, 44-52.
19. Davis, R. C., Buchwald, A. M., & Frankmann, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychol. Monogr.*, 1955, 69, No. 20 (Whole No. 405).
20. Davis, R. C., Lundervold, A., & Miller, J. D. The pattern of somatic response during a repetitive motor task and its modification by visual stimuli. *J. comp. physiol. Psychol.*, 1957, 50, 53-60.
21. Di Mascio, A., Boyd, R. W., & Greenblatt, M. Physiological correlates of tension and antagonism during psychotherapy. A study of "interpersonal physiology." *Psychosom. Med.*, 1957, 19, 99-104.
22. Di Mascio, A., Boyd, R. W., Greenblatt, M., & Solomon, H. C. The psychiatric interview: A sociophysiological study. *Dis. nerv. System*, 1955, 16, 2-7.
23. Dittes, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing statements." *J. abnorm. soc. Psychol.*, 1957, 54, 187-191.
24. Dittes, J. E. Galvanic skin response as a measure of patient's reaction to therapist's permissiveness. *J. abnorm. soc. Psychol.*, 1957, 55, 295-303.
25. Doust, J. W. L., & Schneider, R. A. Studies on the physiology of awareness: The differential influence of color on capillary blood-oxygen saturation. *J. clin. Psychol.*, 1955, 11, 366-370.
26. Doust, J. W. L., & Schneider, R. A. Studies on the physiology of awareness: An oximetrically monitored controlled stress test. *Canad. J. Psychol.*, 1955, 9, 67-78.
27. Duffy, E. The concept of energy mobilization. *Psychol. Rev.*, 1951, 58, 30-40.
28. Duffy, E. The psychological significance of the concept of "arousal" or "activation." *Psychol. Rev.*, 1957, 64, 265-275.
29. Funkenstein, D. H. Nor-epinephrine-like and epinephrine-like substances in relation to human behavior. *J. nerv. ment. Dis.*, 1956, 124, 58-67.
30. Furer, M., & Hardy, J. D. The reaction to pain as determined by the galvanic skin response. *Res. Publ. Ass. nerv. ment. Dis.*, 1950, 29, 72-89.

31. Grace, W. J., Wolf, S., & Wolff, H. G. *The human colon*. New York: Hoeber, 1951.
32. Holmes, T. H., Goodell, H., Wolf, S., & Wolff, H. G. *The nose*. Springfield, Ill. Thomas, 1950.
33. Holmes, T. H., & Wolff, H. G. Life situations, emotions and backache. *Res. Publ. Ass. nerv. ment. Dis.*, 1950, 29, 750-772.
34. Hovland, C. I., & Riesen, A. H. Magnitude of galvanic and vasomotor response as a function of stimulus intensity. *J. gen. Psychol.*, 1940, 23, 103-121.
35. Lacey, J. I. Individual differences in somatic response patterns. *J. comp. physiol. Psychol.*, 1950, 43, 338-350.
36. Lacey, J. I., & Van Lehn, Ruth. Differential emphasis in somatic response to stress: An experimental study. *Psychosom. Med.*, 1952, 14, 71-81.
37. Lacey, J. I., Bateman, Dorothy E., & Van Lehn, Ruth. Autonomic response specificity and Rorschach color responses. *Psychosom. Med.*, 1952, 14, 256-260.
38. Lacey, J. I., Bateman, Dorothy E., & Van Lehn, Ruth. Autonomic response specificity. An experimental study. *Psychosom. Med.*, 1953, 15, 8-21.
39. Lacey, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. N. Y. Acad. Sci.*, 1956, 67, 123-164.
40. Lacey, J. I., & Lacey, Beatrice C. The relationship of resting autonomic activity to motor impulsivity. *Res. Publ. Ass. nerv. ment. Dis.*, 1958, 36, 144-209.
41. Lacey, J. I., & Lacey, Beatrice C. Verification and extension of the principle of autonomic response stereotypy. *Amer. J. Psychol.*, 1958, 71, 50-73.
42. Lasswell, H. D. Verbal references and physiological changes during the psychoanalytic interview: A preliminary communication. *Psychoanalyt. Rev.*, 1935, 22, 1-24.
43. Lasswell, H. D. Certain prognostic changes during trial (psychoanalytic) interviews. *Psychoanalyt. Rev.*, 1936, 23, 241-247.
44. Lindsley, D. B. Emotion. Chapter 14. In Stevens, S. S. (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
45. Malmö, R. B. Symptom mechanisms in psychiatric patients. *Trans. N. Y. Acad. Sci.*, 1956, 18 (Series II), 545-549.
46. Malmö, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, 64, 276-287.
47. Malmö, R. B., Boag, T. J., & Smith, A. A. Physiological study of personal interaction. *Psychosom. Med.*, 1957, 19, 105-119.
48. Malmö, R. B., & Davis, J. F. Physiological gradients as indicators of "arousal" in mirror tracing. *Canad. J. Psychol.*, 1956, 10, 231-238.
49. Malmö, R. B., & Shagass, C. Physiologic studies of symptom mechanisms in psychiatric patients under stress. *Psychosom. Med.*, 1949, 11, 25-29.
50. Malmö, R. B., Shagass, C., & Davis, F. H. Specificity of bodily reactions under stress. A physiological study of somatic mechanisms in psychiatric patients. *Res. Publ. Ass. Res. nerv. ment. Dis.*, 1950, 29, 231-261.
51. Malmö, R. B., Shagass, C., & Davis, F. H. Symptom specificity and bodily reactions during psychiatric interview. *Psychosom. Med.*, 1950, 12, 362-376.
52. Malmö, R. B., Smith, A. A., & Kohlmeier, W. A. Motor manifestation of conflict in interview: A case study. *J. abnorm. soc. Psychol.*, 1956, 52, 268-271.
53. Martin, B. Galvanic skin conductance as a function of successive interviews. *J. clin. Psychol.*, 1956, 12, 91-94.
54. McCurdy, H. G. Consciousness and the galvanometer. *Psychol. Rev.*, 1950, 57, 322-327.
55. Mittelman, B., & Wolff, H. G. Affective states and skin temperature: Experimental study of subjects with "cold hands" and Raynaud's syndrome. *Psychosom. Med.*, 1939, 1, 271-292.
56. Mittelman, B., & Wolff, H. G. Emotions and gastroduodenal function. *Psychosom. Med.*, 1942, 4, 5-61.
57. Mittelman, B., & Wolff, H. G. Emotions and skin temperature: Observations on patients during psychotherapeutic (psychoanalytic) interviews. *Psychosom. Med.*, 1943, 5, 211-231.
58. Mowrer, O. H., Light, D. H., Luria, Zella, & Seleny, Marjorie P. Tension changes during psychotherapy. In O. H. Mowrer (Ed.), *Psychological theory and research*. New York: Ronald Press, 1953. Pp. 546-640.

59. Reiser, M. F., Weiner, H., & Thaler, M. Patterns of object relationships and cardiovascular responsiveness in healthy young adults and patients with peptic ulcer and hypertension. *Psychosom. Med.*, 1957, 19, 498.
60. Rowland, L. W. The somatic effects of stimuli graded in respect to their exciting character. *J. exp. Psychol.*, 1936, 19, 547-560.
61. Ruckmick, C. A. *The psychology of feeling and emotions*. New York: McGraw-Hill, 1936.
62. Schachter, J. Pain, fear, and anger in hypertensives and normotensives: A psychophysiological study. *Psychosom. Med.*, 1957, 19, 17-29.
63. Shagass, C., & Malmö, R. B. Psychodynamic themes and localized muscular tension during psychotherapy. *Psychosom. Med.*, 1954, 16, 295-314.
64. Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
65. Sines, J. O. Conflict-related stimuli as elicitors of selected physiological responses. *J. proj. Tech.*, 1957, 21, 194-198.
66. Spence, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, 13, 131-141.
67. Stennett, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, 54, 54-61.
68. Terry, R. A. Autonomic balance and temperament. *J. comp. physiol. Psychol.*, 1953, 46, 454-460.
69. Thetford, W. N. An objective measurement of frustration tolerance in evaluating psychotherapy. In W. Wolff & J. A. Precker (Eds.), *Success in psychotherapy*. New York: Grune & Stratton, 1952. Pp. 26-62.
70. Vogel, W., Baker, R. W., & Lazarus, R. S. The role of motivation in psychological stress. *J. abnorm. soc. Psychol.*, 1958, 56, 105-112.
71. Wenger, M. A. A study of physiological factors: The autonomic nervous system and the skeletal musculature. *Human Biology*, 1942, 14, 69-84.
72. Wenger, M. A. Studies of autonomic balance in Army Air Forces Personnel. *Compar. Psychol. Monogr.*, 1948, 19, No. 4 (Serial Number 101).
73. Wenger, M. A., & Gilchrist, J. C. A comparison of two indices of palmar sweating. *J. exp. Psychol.*, 1948, 38, 757-761.
74. Wolf, S., & Wolff, H. G. *Human gastric function*. New York: Oxford University Press, 1947.
75. Woodworth, R. S., & Schlosberg, H. *Experimental psychology*. New York: Holt, 1954.

Discussion of Papers by Saslow & Matarazzo, and Lacey

MILTON GREENBLATT, M.D.

I have been asked to discuss two excellent papers—one by Saslow and Matarazzo having to do with intensive analysis of the parameters and qualities of a highly specific interview procedure—the interaction chronograph method of Chapple; the other paper by Lacey concerning the use of physiological data to evaluate psychotherapy. These two papers have little in common. Saslow's paper has not a word about physiology although Saslow himself has been extensively trained in physiology. It is specific in its focus as it examines only one interview procedure and is concerned primarily with the use of this procedure as a scientific instrument. It is closely written and closely reasoned, passing from one reliability question to another in systematic fashion. Although the question of validity is raised, relatively little validity data are given and the paper, too, has relatively little concern with *content* of the interview. Lacey's paper, on the other hand, is broad in scope and ranges sensitively over the whole field of interview psychophysiology. It examines critically many different approaches, techniques and experimental models; it discusses content, affect, conflict, transference and counter-transference, and all the other elements that constitute the soul of psychotherapy. He attempts to put the data critically within a systematic pattern of thought developed from his own experience, and even searches for new integrative concepts to help us along in psychophysiology.

It is interesting that Saslow the *psychiatrist* works in a relatively impersonal way with *time* functions of an interview almost exclusively, eliminating, as I have said, content analysis almost entirely;

while Lacey, primarily the *psychologist*, has tried to come to grips with the refinements of the human aspects of personal interchange.

In the Saslow and Matarazzo paper, the focus is on the idea that the interview is "unreliable" as an instrument of measurement in its free form and therefore requires standardization. However, they are concerned that the standardization be maintained within limits to allow for preservation of the dynamic richness of the interview itself.

The *time* relations of interactions are selected as a significant and easily measurable dimension. Immediately we ask, and they ask, To what extent will this quantitative aspect of analysis allow one to make inferences as to emotion, transference, resolution of conflict, etc.?

In defense of this time measure, Chapple indicates that it is a common experience that people we *like* may say the same things as people we *don't like*, but that the *timing* is different. If this were true, the method would perhaps be worth more than I think it is. So many elements other than timing seem to play a role in the interview; take, for example, inflection. There is the interesting story of the tailor who had a sign in his window that said, "My name is Fink—What do you think—I sell *suits* for nothing." But when people came in to claim his suits, he said, "My name is Fink. What do you think? I sell suits for *nothing*!?" Here, timing may be different, but I suspect the inflection is the major thing.

To many of us, the psychotherapeutic interview has *therapeutic* value insofar as it permits *freedom* of expression and

behavior of the patient, and we might also say the psychiatrist. If this concept is correct, the *natural variations* in mood, activity, etc.—yes, and timing, too, become the *essential* elements in this transaction worthy of study. If it is free communication that is wanted, we ask, Why labor so hard to standardize the situation? Cronbach has defined two methods of psychological research: the “experimental” and “correlative”—the latter which takes advantage of variations in the experiments of nature may be more applicable to the interview situation than the former.

I have some quarrel with the choice of designations for various functions abstracted in the Chapple interaction analysis. For example, Table 1, “Patient’s Adjustment” is defined as “the durations of the patient’s interruptions minus the durations of his failures to respond, divided by Patient’s Units.” The word “adjustment” is a highly charged and value-laden term, especially to psychologists and psychiatrists. Are we to assume that adjustment is better if interruptions are longer, worse if interruptions are shorter?

The patient’s “dominance” is defined as “the number of times (out of 12) in Period 4 that the patient ‘talked down’ the interviewer minus the number of times the interviewer ‘talked down’ the patient, divided by the number of Patient’s Units in the Period.” Again the patient would appear to be more dominant if he talked down the interviewer, less dominant if he did not. Politeness and good breeding might imply that if one is repeatedly interrupted by an interviewer, one should possibly let the poor fellow have his say—he is probably too rigid and lacking in insight to understand anyway. If this is a reasonable interpretation, one could suggest instead of “dominance” the word “boorishness.”

The patient’s “synchronization” is defined as “the number of times the patient

either interrupted or failed to respond to the interviewer, divided by the number of Patient’s Units.” I am at a loss to interpret its meaning. Perhaps it is my unfamiliarity with this particular concept. Do the authors feel that terms with a less derivative quality could be used? I feel this matter of terminology is not merely a semantic quibble.

The authors have ably demonstrated how successfully it is possible to standardize the interview. It is truly extraordinary, as they indicate, that the coefficient of correlation for “tempo and activity,” for instance, went from .53 and .79 to .93 and .95 after two interviewers had put their heads together. They also indicate how subtly two members of a dyad influence each other in terms of the duration of utterances. In over eight conversations there is a rank order rho for this function of .83 which is very striking.

The subtle but marked effect noted here of mutual influence of interacting parties was transparent to me in an unusual psychotherapeutic experience wherein five psychiatrists interviewed one psychotic patient at intervals during most of a year. The patient was psychotic; on and off drugs; and her behavior varied over a wide range. When she was quiet and depressed, the group of psychiatrists slowed down and appeared retarded. When facetious, they tended to joke; and when hostile, they became snippy with each other. Tone and mood rose and fell together, both for the patient and her five psychiatrists. This mood concordance is highly worthy of study, and perhaps it correlates with the time dimension noted. (One is reminded also of the story of Rioch’s patient who was depressed on all occasions but six—these were the ones when Rioch was under the influence of benzedrine. Obviously much work has to be done here.)

To the remarkable work of Sidowski, who reinforces plurals by a blinking light,

and Greenspoon, who achieved similar results by the therapeutic "umm-hmm," should be added the work of William Verplanck who was able to push conversational speeds up and down almost at will and towards desired subject material by conversational reinforcements, probably unknown to the patient. This type of "hidden conversational sell"—to paraphrase the currently popular psychological phenomenon—is an excellent demonstration of the conditioning probably basic to all psychotherapy, and when more carefully studied may even help to increase technical proficiency in our use of the interview tool.

Standardization of the interview situation is, of course, the most troublesome question for the psychiatrist who has to be convinced that compressing the interview into a mold still allows psychotherapy to go on and, after all, this is a conference on *psychotherapy*, not merely on the *interview*. Consider the strictures put on the interviewer: he should be non-directive, open-ended, verbal only (can't use head-nods), restrict himself to 5-second utterances, respond in less than one second; if interrupted, continue for two seconds so as to make scoring easier, etc. If subject does not respond, then wait 15 seconds and speak again for 5 seconds. In period 2, fail to respond 12 times. In period 4, interrupt 12 times for 5 seconds, each interruption after 3 seconds of patient talk. Even smoking is interdicted (in later studies). Does the interview become a shadow of the natural one by so much sacrifice to the goal of reliability?

May we also ask, What becomes of the interviewer himself after he has been treated in this fashion? Does he feel that the medium is suitable for true therapeutic work? To what extent is he comfortable in his role? Do the authors feel that the average well-trained psychiatrists could learn to accommodate to these conditions easily or well?

The large sacrifice to the God of reliability has evidently paid off. We learn that observer reliability is high; that a given interviewer's reliability is high for different patients, that it is possible for him to be well trained in silent periods, duration of utterance, failure to respond, interruptions, etc., etc. Two interviewers on the same subject agree well on the length of their interviews. There is relative and absolute reliability of two interviewers on many different variables. There is considerable stability in re-interviewing one and five weeks apart, and even relatively satisfactory agreement on many items after eight months although at this time some wandering from the path of reliability is noticed.

It is interesting that stability is greatest on re-interviewing for those subjects who have had few psychotherapeutic hours (2.3 hours average), and less for those who have had more psychotherapeutic hours (7.2). It would appear that *five hours difference* in psychotherapy over an eight-month period is sufficient to change the reliability significantly. We ask whether these changes would be greater if deep transference or therapeutic effect occurred, for then the strain on both the subject and the doctor to respond to his inner problem would be great, as would the tendency to break beyond the bounds of prescribed technique.

In this comprehensive and detailed study by Saslow and Matarazzo, the subject of validity is touched upon only slightly. However, the authors have apparently conducted investigations relating interview parameters to psychological tests, content analysis, factor analysis, dynamic grouping, ward behavior, and Bales interaction analysis. We look forward very eagerly to these results in order to test our own assumptions as to the possible value of this interview instrument. Perhaps these studies will clarify the question as to whether the *standard-*

ized or the *free interview* is likely to be more scientifically or practically rewarding in the long run. Whatever the results, if they have been given the Saslow and Matarazzo touch of thoroughness, carefulness, and sophisticated research methodology, we shall all be indebted for a real public service.

While I have a conventional role as a discussant of the Saslow and Matarazzo paper, I am at a loss as to what to do with Lacey's paper. You see he has fully discussed the literature on physiology in the interview in an extensive and carefully-reasoned analysis. It would be overdoing it to discuss this discussion. After some rumination I have decided to look upon it as a simple case of role reversal. Inasmuch as our laboratory has been studying psychophysiology of psychotherapy for some time, clearly, I might find a useful role in presenting some of this material which could then be a suitable basis and background for Lacey's discussion. Thus everything would be in order except that the time sequence would be backward. However, the complexities of our problems today are such that a little time rearrangement will hurt no one. In fact, it might provide us with just one more interesting intellectual challenge.

First let us consider the influence of the experimental situation on the psychotherapeutic process. This is a topic which Lacey has not mentioned, yet, it is vital to those of us who are interested in the effectiveness of our psychotherapy. Obviously if recording of the subject and, in some instances at least, of the therapist interfered drastically with the transactions and made therapeutic advances by the patient impossible, we would have little interest in physiological measurements. The paper by Watson and Kanter (8) from our laboratory examines this question critically based on doctor-patient relationship in which both parties were

studied psychophysiological for 44 interviews. The patient was a 34-year-old Ph.D., father of five children who himself was interested in research. He was taken off the waiting list of the Out-Patient Clinic and taken immediately into therapy because he was interested in cooperating. However, at his first interview he expressed anxiety about being observed and sighed with relief when the electrodes were attached to the therapist, saying "You're on the spot, too." He asked numerous questions about the nature of the research and was given honest answers. He never saw the polygraph operators who were in another room, but saw only the technician who applied the electrodes, and the therapist—members of the research team.

To make a long story short, the research situation became a transference problem in which the patient soon resisted giving information of an emotional, personal type because he felt that this was not true therapy. Thereupon, he tried in various ways to break off the contract, to involve the therapist in interchanges outside the observation room—even calling him at night and in his private office—but the therapist stood his ground. He then began to play the game of harboring a very serious and damaging secret about the therapist which could not be told in this situation because of the observers. There was also a third party that might be hurt, he hinted. It was not until the 16th, 17th and 18th interviews that the climax of the working through of this problem occurred. It consisted of a direct discussion of the relationship, its hostile and negative aspects, derogatory remarks about the therapist, and culminated in revelation of the personal situation he had been hinting about—actually a transitory flirtation of some years past between the therapist and the third party known to the patient. The flows of tears at about this time cleared the air and convinced the patient

that he would not be attacked by the therapist as he expected to be treated by authority figures.

This is quoted in some detail to show that the interview situation to the patient has both *reality* and *transference* aspects but that true psychotherapeutic work can go on despite the complications of recording devices. I might say parenthetically that this patient improved greatly as indicated by projection tests and by the judgments of all the observers and his own testimony.

The therapist indicated his reaction to the research and observational nature of the situation, to wit: (a) he had doubts about his professional adequacy as a therapist, and, (b) concern about exposing his personality. "I felt moderately tense, particularly at the beginning and very end of the hour, and somewhat artificially immobile as a reaction to the anxiety. There was some fluctuation in my awareness of tension during the course of the hour." By the second hour he states, "I felt less tense," and he was able to indicate to the patient that the latter's intense distress had better be worked out as a neurotic affect. By the fourth interview the therapist was much more comfortable as it became clear that a therapeutic relationship could be and was being established. At the termination of therapy, the psychiatrist stated as follows: "As my activity increased with the unfolding of the dynamic conflicts, I felt I was functioning as I do in the clinic or in private practice, and that the research situation no longer influenced my behavior as a therapist."

The upshot of the whole study is put in these words: "The addition of observation, tape recording, and other investigative procedures to the therapeutic situation introduces psychodynamic elements of the same order as the specific attributes of the patient, the therapist, and the therapeutic environment . . . *the research aspects are not an element of*

greater significance than other transference phenomena which influence therapeutic progress." (Author's italics)

Thus we can be reassured that observation will not "kill" psychotherapy but it certainly can complicate it, for the meaning of the total situation to either party cannot be easily understood until a fair section of the psychodynamics of each individual is understood.

To what extent can physiological recording reflect emotional activity of individuals during psychotherapy? Figure 1 (6) shows the temperature curves for this subject for interviews judged "high emotional participation" and "low emotional participation." Note that the interviews with "high emotional participation" started at low skin temperature levels and the curves show considerable lability or fluctuation. On the other hand, interviews with "low emotional participation" are characterized by high average skin temperature levels and little fluctuation.

Perhaps the most striking example of "skin language" is illustrated by Figure 2 which is the finger temperature curve of a student nurse who was one of a group of "normal" subjects undergoing a series of stress tests in order to define the parameters of this population. Note the rapid drop of skin temperature when the psychiatrist enters the room, and the tendency to rise when he leaves. The anxiety material concerned an illegitimate pregnancy suffered by this young lady that she had been anticipating talking about in this session. It is rare to have the combination of high reactivity and immediate focusing on an emotional problem of such magnitude, however, this does show how extraordinarily responsive an individual *can* be in one somatic area. As Lacey has pointed out so well, other members of our series showed no change at all and were totally unreactive in the skin temperature dimension although the experimental situation was similar. There is no question

Temperature Curves for Two Levels of Emotional Participation
(First 6 Interviews)

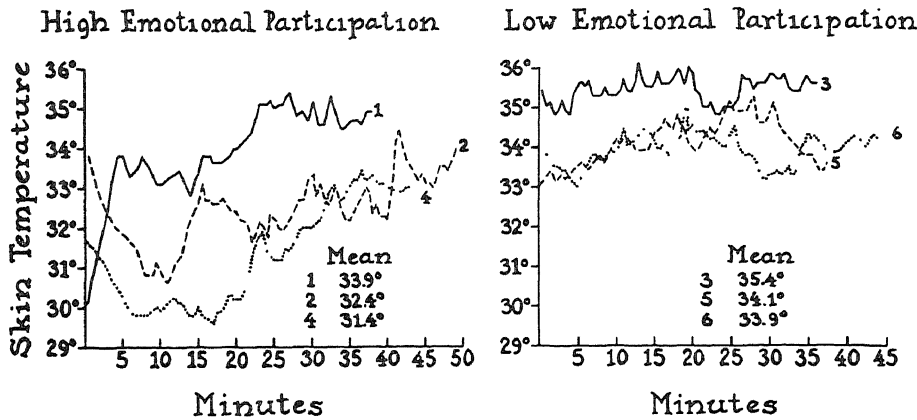


FIGURE 1
Temperature curves for the first six interviews of one subject during psychotherapy. Interviews high in emotional participation (1, 2, 4) are compared with interviews low in emotional participation (3, 6, 5). Interviews high in emotional participation are characterized by relatively low initial temperature level and wide fluctuations, and interviews low in emotional participation are characterized by relatively high initial temperatures and more stable temperature curves throughout the interview period

CHANGES IN FINGER SKIN TEMPERATURE IN A "REACTIVE" SUBJECT DURING AN EXPERIMENTAL SESSION

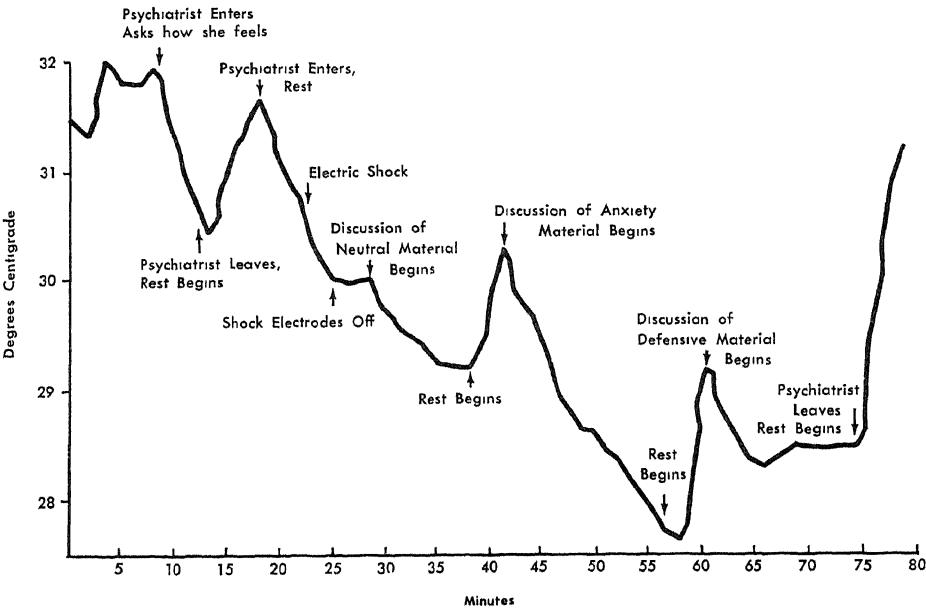


FIGURE 2
Skin temperature reactions of young affiliating nurse during interview with a psychiatrist.

of the importance of *ego involvement in touching off physiological change*.

In another paper as yet unpublished (4), the changes in skin temperature in the first case mentioned above were analyzed in relation to ego functioning as reflected by the psychodynamics of the case.

The relationship of mastery or non-mastery of anxiety to level and changes of finger skin temperature were studied. The conclusions in this study are framed as follows: the fluctuations in skin temperature tend to mirror the shift in the balance between *ego resources* and *internal or external stresses* including the stress of *facing one's self in psychotherapy*. A drop in finger skin temperature accompanies the ego's actively working at a problem in therapy, whereas a rise is associated with mastery through defensive or adaptive methods. (More specifically, it was noted that level of free floating anxiety as judged on the clinical data correlates negatively with level of skin temperature. The expressions of depressive or affectionate feeling were characterized by high, stable skin temperature; the expressions of anxiety and hostility were characterized by lower and more fluctuating skin temperature. The direction of skin temperature change was often upward with the first set of conditions, and downward in the second.)

The temperature findings are to some extent paralleled by our studies of fluctuations of palmar skin potential in response to stress. Let me quote from a study of Learmonth and Ackerly (5), as yet unpublished.

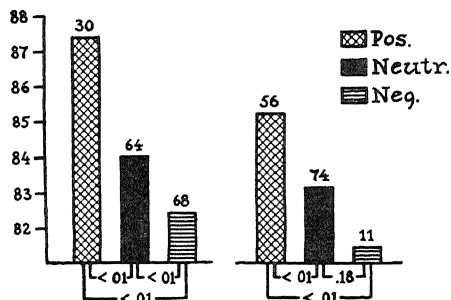
"The galvanic skin response of 20 student nurses during periods of basal rest, sentence completion stress, interview stress, and physical stress were analyzed to determine [the correlates of]¹ the in-

crease in fluctuation of palmar potential in response to stress. Results were correlated with rankings of the same subjects on 22 personality variables obtained from the MMPI and Rorschach given at a later date. The following inferences were suggested.

"1. The increase of fluctuation of palmar potential in response to stress is negatively correlated with a group of personality features which have in common the element of expressivity

"2. The increase of fluctuation of palmar potential in response to stress is correlated positively with a group of personality features which have in common the restraint and curtailment of unpleasant or prohibited feelings and actions."

Patient's Mean Pulse with Bales' Interaction Units



Patient's Units Doctor's Units

L.W. d'34, Diagn. Depressed 1-22-1953

FIGURE 3

The heart rate of this patient during psychotherapy tended to be higher during positive expressions (as defined by Bales' interaction categories) than during neutral expressions, and lowest during negative statements. Positive, neutral and negative statements by the physician were associated with a similar pattern of heart rate of the patient during this interview. The number of positive, neutral and negative expressions of patient or of physician are indicated above each column. The significance of the differences is also indicated.

1. Author's insert.

Changes in skin temperature and in palmar potential are but two examples of sensitive fluctuations that may occur in physiological parameters during psychotherapy. Perhaps one of our definitive contributions in this area is our recognition of how sensitively the *heart rate* of individuals may register psychological changes. Increasing experience with this problem led us to have enormous respect for the moment-to-moment changes in heart rate, and to realize that clinical methods that utilize long intervals like 15 seconds as a basis for estimates of rate are crude and unrefined.

In any interview one is likely to see broad changes in heart rate as well as brief ups and downs. We got some evidence of the significance of these relatively minor fluctuations when we cor-

related the Bales Interaction units with the heart rate. Figure 3 (3) shows that when the patient is making a positive statement his pulse is higher than with a neutral statement, and that it is higher again than with a negative statement. This finding surprised us but has been shown in several more instances. Bales' analysis of the patient's heart rate to the doctor's expressions shows a similar pattern. The patient, therefore, within the setting responds similarly to the doctor's affective utterances as he does to his own. This finding shows strikingly the transactional nature of the interview that has been so underscored by Lacey. When large portions of the interview were judged as to whether the patient showed positive, neutral, or negative attitudes, or transference to the doctor, some simple

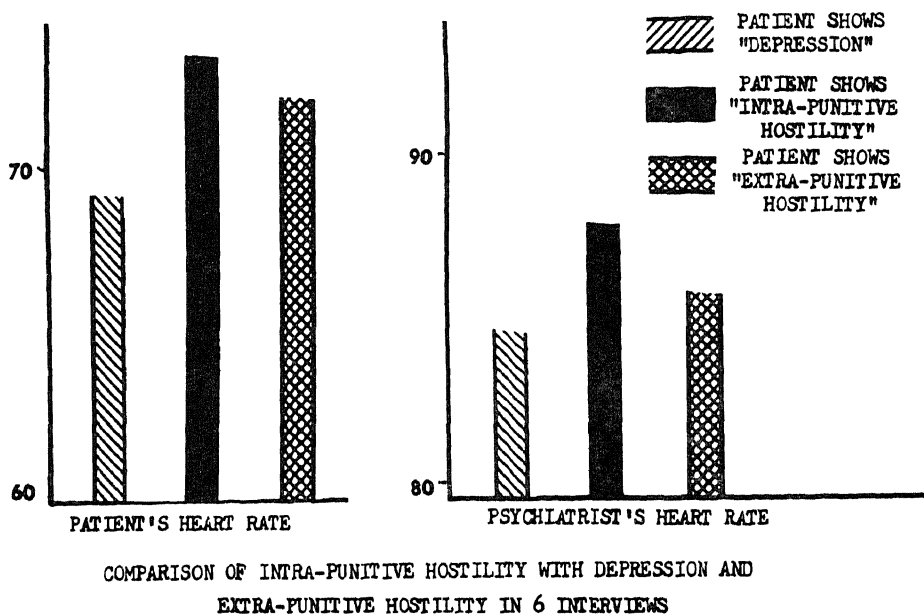


FIGURE 4

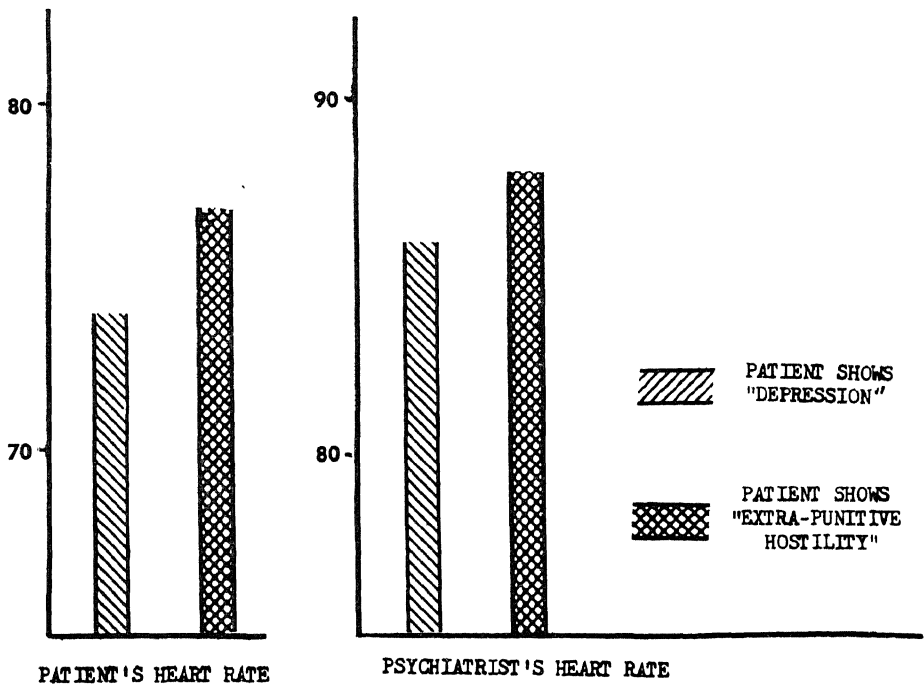
Note that both patient's and psychiatrist's heart rate tend to be highest when the patient shows intra-punitive hostility, lower when the patient shows extra-punitive hostility, and lowest when the patient shows depression even though the average level of heart rate is different for the two parties.

and significant differences did occur in heart rate level and in heart lability.

The doctor's physiological activity in the interview must not be neglected. Studies of different doctor-patient diads have shown us that the doctor is quite as reactive as the patient. Any notion that the doctor is a simple reflecting screen has been laid low by these observations. I would like to refer to the study of Coleman, Solomon and myself on physiological evidence of rapport (1) to give underpinnings to Lacey's discussion. I could also perhaps correct a possible misinterpretation of the study by saying that indication of the observed affect which is very brief must not be confounded with the temporal aspects

of that affect. The affects were not instantaneous although the identification of affect was made as one brief code. The simultaneous rate of the patient and the doctor's EKG is the object of analysis. Figures 4, 5 and 6 suggest physiological rapport at least for *some* of the emotions experienced by the patient. It is further worth noting that the rapport phenomenon was most striking when the doctor was "actively listening" and less striking when he was distracted or "not listening."

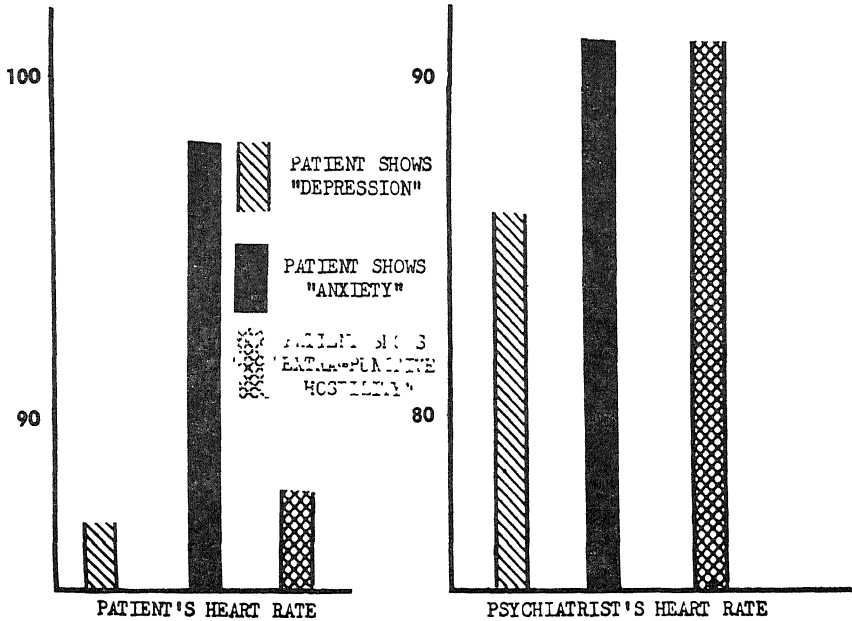
For the interview as a whole, physiologically meaningful findings of an interactional nature were obtained. Using again Bales' categories of "tension release," "neutral affect," and "tension,"



COMPARISON OF DEPRESSION WITH EXTRA-PUNITIVE HOSTILITY IN 19 INTERVIEWS

FIGURE 5

Note that both patient's and psychiatrist's heart rate are higher for extra-punitive hostility than for depression even though the average level of heart rate is different for the two parties.



COMPARISON OF ANXIETY WITH DEPRESSION AND EXTRA-PUNITIVE HOSTILITY IN 2 INTERVIEWS

FIGURE 6

Note that the patient's heart rate goes much higher during his anxiety than does the psychiatrist's heart rate.

we find that the frequency of these units in Bales' analysis for a given interview correlates with the heart rate and lability.²

Interviews with many "tension" scores are inclined to show higher heart rates for both patient and therapist than interviews with few "tension" scores. Also "high tension" interviews tend to show low lability of heart for both patient and psychotherapist. On the other hand, interviews with high frequency of "antagonism" units were correlated with a higher therapist heart rate and a lower patient heart rate. These findings can be interpreted in many ways. They show that a

relationship exists between the physiological activity of patient and therapist, sometimes positive, sometimes negative, sometimes random (2).

To complicate the situation a little more and underscore another one of Lacey's points which is that the patient may be reacting to the whole situation, we can add that not merely is there physiological response to himself, to the physician, to the interpersonal atmosphere, to the physical and social context of the interview, but the whole therapeutic process itself may be subject to variations imposed by climatic shifts. Soon cosmic forces and outer space considerations will come into the picture. However, at the moment we have mundane data (7) concerning the relationship of air temperature, barometric pressure and humidity to the doctor's and

2. In Bales' scheme, the patient shows "tension release" when he jokes, laughs, shows satisfaction; he shows "tension" when he asks for help or withdraws out of the field; and he is "neutral" in his affect when he gives suggestion, opinion or orientation or asks for suggestion, opinion or orientation.

patient's physiology that are highly relevant.

Following the 44 interviews of which I have spoken, our researchers obtained climatic data from nearby Logan Airport relevant to the conditions prevailing at the time of each interview. These interviews were started in July and terminated in February. The therapeutic room was air-conditioned, the temperature being around 72 degrees. There was no control of barometric pressure or humidity in the experimental room. We have data relative to therapist and patient heart rate and lability, patient's skin temperature and lability during the summer months only, and patient's average rate of speaking. The correlations were as follows:

1. The patient's finger-skin temperature level and lability (data available

only for the summer) were related to each of the climatic variables—*level* with air temperature $+.73$, relative humidity $-.82$, barometric pressure $-.49$; *lability* with air temperature $-.66$, relative humidity $+.66$, barometric pressure $+.80$.

2. The therapist's heart rate correlated negatively with the relative humidity ($- .51$), whereas the patient's heart measures showed no significant correlation with the climate.

3. The patient's rate of speaking correlated negatively with atmospheric pressure ($- .47$).

This study suggests that in psychophysiological experiments, particularly in studies of psychotherapy where level or change in activity is being measured, climatic factors (atmospheric pressure as well as temperature and humidity) should

CORRELATIONS (RHOS) OF SOCIAL INTERACTION CATEGORIES
AND PHYSIOLOGIC ACTIVITY (HEART RATE) DURING PSYCHOTHERAPY

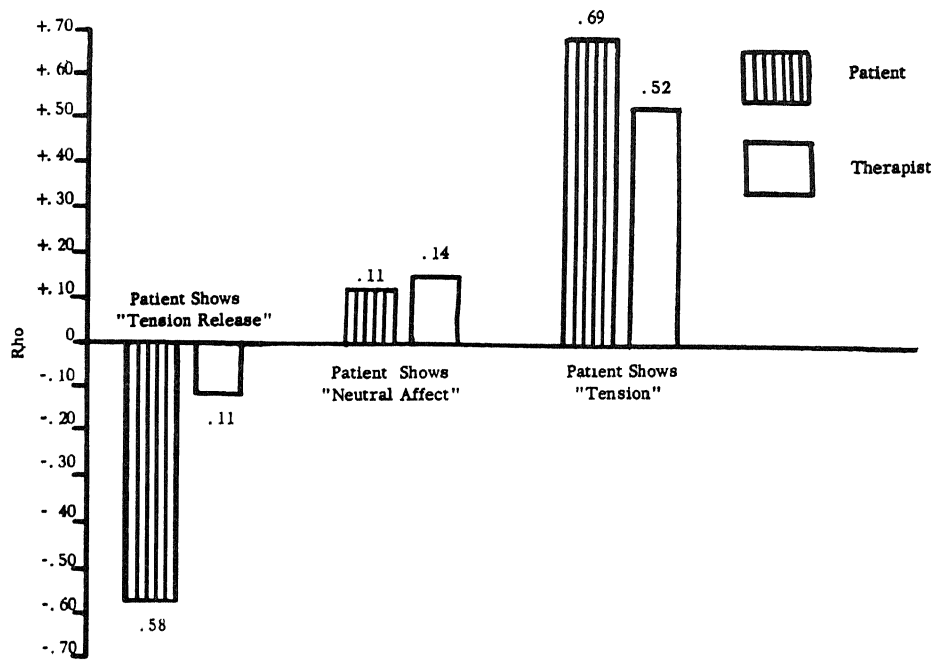


FIGURE 7

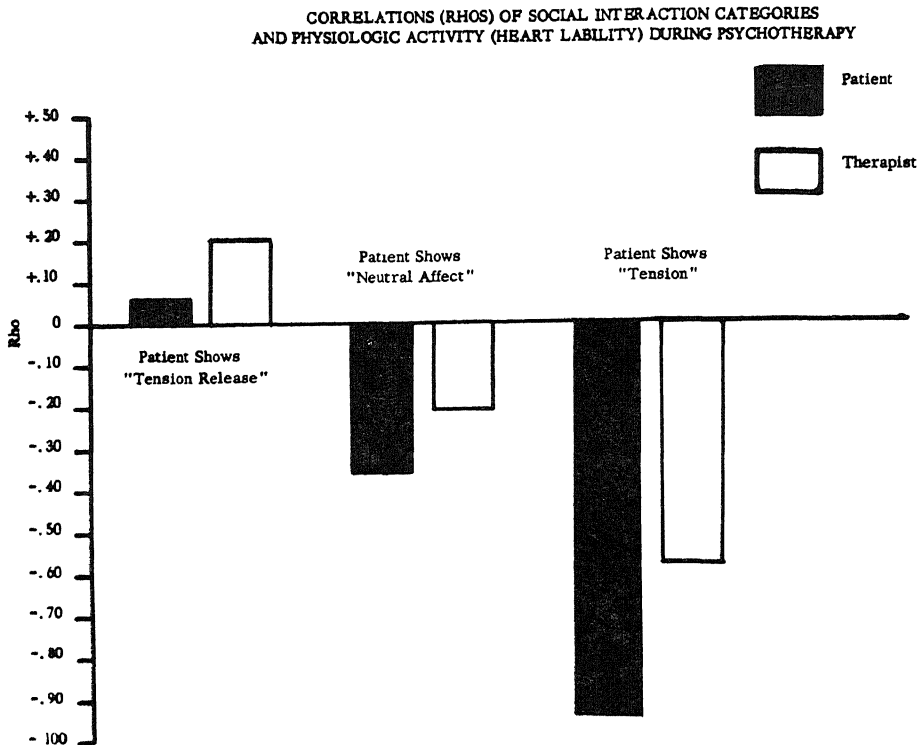


FIGURE 8

be controlled or taken into account, especially since the two parties in the therapeutic relationship may respond differently to climatic variations.

Perhaps I have mentioned enough of our work to underscore Lacey's wise emphasis regarding the need for critical judgment and reserve in the interpretation of physiological data obtained in that "welter of complexity"—the psychotherapeutic interview.

REFERENCES

1. Coleman, R., Greenblatt, M., & Solomon, H. C. Physiological evidence of rapport during psychotherapeutic interviews. *Dis. Nerv. Sys.*, 1956, 17, 71-77.
2. DiMascio, A., Boyd, R. W., & Greenblatt, M. Physiological correlates of tension and antagonism during psychotherapy. *Psychosom. Med.*, 1957, 19, 99-104.
3. DiMascio, A., Boyd, R. W., Greenblatt, M., & Solomon, H. C. The psychiatric interview. *Dis. Nerv. Sys.*, 1955, 16, 4-9.
4. Kanter, S. S., DiMascio, A., & Watson, P. D. Some aspects of ego functioning as reflected in skin temperature change. *J. Nerv. & Ment. Dis.*, in press.
5. Learmonth, G., & Ackerly, W. Palmar skin potential fluctuations in response to stress. Unpublished manuscript from Massachusetts Mental Health Center Research Laboratories, 1958.
6. Massachusetts Mental Health Center. Psychophysiological studies of the psychotherapeutic process. USPHS Project M-354(C3), Progress Report, February, 1, 1956.
7. Watson, P. D., DiMascio, A., Kanter, S. S., Suter, E., & Greenblatt, M. A note on the influence of climatic factors on psychophysiological investigations. *Psychosom. Med.*, 1957, 19, 419-423.
8. Watson, P. D., & Kanter, S. S. Some influences of an experimental situation on the psychotherapeutic process. *Psychosom. Med.*, 1956, 18, 457-470.

Methods for Assessment of Change (B)

DR. BUTLER: I think it is very impressive that some regularities are found in a standard situation. It seems to me that among the greatest regularities that we find are the interaction patterns of people. This is observationally true, also. By a coincidence, roughly ten years ago we had an inadvertent control—the kind that Dr. Robbins and Wallerstein referred to—in which we had two therapists of quite different persuasion working in collaboration on one client. Working in collaboration they had pairs of interviews, four or five minutes apart. The client showed quite different patterns of interaction, very noticeable, and quite different patterns of content with these two therapists.

We ran a chronograph on this material—not on the visual material, but on the sound material—and one very significant thing I think that we found was that upon factor analysis of the data, it appears that the same chronograph factors appear for both therapists, although the interviews perceptually speaking, rather than chronographically speaking, had different content and in the interviews seemingly the individuals showed different styles of approach to the therapy so that the personality almost seemed different. The same interaction factors emerged on the chronograph for both therapists. They emerged at different times in therapy, however.

I think this is a remarkable finding showing the constancy in chronograph terms of the kinds of patterns that are found, and it does emphasize that tem-

poral factors as well as inflections and so forth somehow seem important in interaction between therapist and client in a rather fundamental way. I think it casts some light on this question that Dr. Greenblatt raised, "What about inflection?" No doubt inflection is important, but it shows also that these quantitative factors somehow do bring out significant information.

DR. BORDIN: My question is very relevant to what Jack Butler said. I am worried that if you get too much constancy, this seems to remove it from something that we are interested in from the point of view of psychotherapy.

DR. MATARAZZO: I think this is an extremely important point and one which we certainly should discuss a little more. The constancy is not universal in this sense. The constancy that we found in our studies is *not* for any given individual *under all conditions*, but rather any individual *exposed to the same kind of standardized interview* will emit essentially the same pattern on retest whether the retest be by another interviewer or the same interviewer eight months later. This fact of the stimulus conditions being identical is important.

Secondly, that same client or patient, if retested by a different interview method will not give the same pattern. For example, if I were to interview Patient 1 using my own *free* interviewing style and Carl Rogers were to interview that same patient by use of his own *free* style, the pattern he would get from the patient would be very, very different from the pattern I would get. That is, a correla-

* Abridged. See Editors' note, page 49.

tion between what he got and I got would be of the order of zero. However, if he and I decided just to standardize a few of the interviewing rules in advance, the agreement between him and me would be of the order of correlations of .90 on a number of interaction variables. So the patient is *not* a constant organism, in this sense. Rather the patient's behavior under identical retest conditions will be the same as his earlier interview behavior.

A second point is important here. No two patients have interview patterns that are alike. I think Dr. Saslow indicated earlier that one patient might talk with him 24 times, and a second patient *under the same conditions* of the interview will talk 124 times. Thus huge individual differences under identical interview-stimulus conditions are also possible. The important thing is that if I reinterviewed Dr. Saslow's Patient 1, he would talk either 24 or 25 times with me, if I used the same standardized technique. Similarly, if I reinterviewed the second patient, using the standardized method, he would talk 124 or 125 times with me. This has an extremely important bearing on research in psychotherapy. If we are interested in measuring change in psychotherapy, and we have found in the past that assessment techniques like the Rorschach test and similar techniques do not reflect the change in behavior which the therapist and the others think have occurred, we now have an apparently useful technique which will allow us to assess these changes in behavior.

I will give an example. An individual at the start of therapy might talk under the conditions of the standardized interview a total of, say, 88 times, with each unit of 15 seconds duration. At the end of therapy our test-retest reliability studies show this same patient will talk 88 times plus or minus a few units, if nothing has occurred in psychotherapy to change his behavior. If, instead, we

find he talks 42 times and the duration of each utterance is 47 seconds, then we know a change has taken place. Since we have done control studies, we can attribute the change, in all probability, to this intervening variable, namely, psychotherapy.

DR. BUTLER: I did not express myself clearly, but our interview situation was an unstandardized situation. This is what I think was so remarkable. Two therapists. Second, it was only sound recording instead of the complete material you had. Third, the patterns were different, but the factors were the same.

DR. MATARAZZO: Unless I am mistaken this is a study by Lundy which was done at the University of Chicago. The interesting thing is that the factor analysis of the interview interaction patterns of the two therapists at Chicago gave essentially the same five interview factors Working in St. Louis, and factor-analyzing the material of our single interviewer-therapist, in this case George Saslow, we got three of these same five factors. We believe the reason we did not get the other two was that we were not using all of the interview measures Lundy was, and the remaining two factors Lundy obtained were relevant to these absent variables in our study.

DR. ROTTER: This is on the same topic. I think it is a very exciting technique to study some of the aspects of the interaction of two people in an interview situation. What bothers me a little is that you keep on talking about an N of 20 subjects and two interviewers, or three interviewers, and so on. It is quite possible that the personal tempos of these interviewers is quite similar. What I am interested in is 20 interviewers. Some of these statements about the constancy of the subject may entirely be artifacts of some accidental similarities of your interviewers. Do you have any evidence on this?

DR. MATARAZZO: That is a good point. We have *purposely* made our interviewers alike. They *were not* alike at the beginning. After we had them learn the method they became like one. In other words, the glucose tolerance test used in St. Louis was like the one used in Boston.

DR. ROTTER: This is something which would be nice to do experimentally perhaps with selection in terms of initial personal tempos on an untrained sample, and then follow through with your training at the various levels.

DR. SASLOW: We have thought of such an experiment using a group of new residents, maybe at least half a dozen. As an indication of a possible answer, take the observations that were made at the Massachusetts General Hospital when Chapple was first using this method. There were two psychoanalysts who were interviewing. These two men had been trained in very similar ways. They were of comparable experience and professional stature, held the same personality theory, and they themselves believed that they were carrying out identical interviews, with military personnel as subjects, in a particular series of observations. They were very surprised when the simplest measurements made by Chapple demonstrated conclusively that they elicited very different patterns of behavior from each subject interviewed by them.

Their willingness to agree to standardize their behaviors in just a few ways, for example, the free give-and-take of our Period I behavior, with no interruption, with prompt response, etc., immediately increased the agreement between the two of them to a very noticeable degree, whereas previously the disagreements were obvious despite the fact that they were not perceptible from the type-scripts, or remembered content or subject-interview behavior.

Some of the points Dr. Greenblatt raised in his discussion paper are very important. Labels that Chapple has used for variables are built into an instrument which we cannot change. We would like to experiment with variables that are simpler than his and with different ones. He has used terms which have both operational and colloquial meanings. The term "adjustment," for example, has an operational definition in terms of the interaction chronograph, but the colloquial meaning gets in the way; the same is true of "dominance." We have used all the variables in their operational sense only.

In my introductory remarks, I stated that we have not been working with psychotherapy as a process. I would never dream of using such a procedure, as it stands, as psychotherapy. On the other hand, there are tempting implications that analogous procedures might be tried as part of psychotherapy and that it might be possible to appraise their effects objectively. It is also clear that a pragmatic temperament has cautioned me against premature use of physiological methods in the study of psychotherapy, and I am waiting for Lacey and others to work out some of the obvious complexities. As a matter of fact, Lacey and we have become interested in seeing whether we can study some subjects both by the methods he has found will reveal bursts of autonomic system activity, and by the interaction chronograph method. Responses defined in these two very dissimilar ways may well have interesting relationships.

I would share your discomfort in trying to conduct a series of interviews exactly like this, and it would never occur to one to do that. However, if one objects to cautious standardization of even a partial degree, one is raising a rather different kind of question. You can object to the psychological test situation as not leaving the subject free to

walk out. Standardization it seems to me is justified in terms of what the researcher is after, and his awareness of what he can't get by the particular standardizations he introduces. I think one should point out what the standardization causes you to lose as well as what it allows you to gain.

DR. MATARAZZO: We were interested purely for reliability purposes (i.e., habituation possibilities) in what would happen if the same standardized interview were repeated ten days sequentially, or during ten interviews separated by some other interval of time. That is, we wondered if you would get the same identical pattern elicited from the subject each and every day. What we found was no, you did not, because as the interviews progressed this high reliability which occurred between interviews 1 and 2 no longer occurred. The further along you went in these interviews the more and more change you got. I was interested in this purely from the point of view of using the same technique in time, just to see what kinds of responses we would get. Would we get this high stability in patient interaction behavior from Days 1, 2, 3 and 4? Where we did initially get high stability from Day 1 to Day 2, as interviewing went on (and observers on the other side of the one way screen who were watching me said changes were taking place,—minimal, of course, in such few "psychotherapy" sessions) the interaction chronograph patterns reflected the fact that this man at the end of ten interviews was very different than he was at the beginning. That is, the reliabilities of the latter sessions with Day 1 were very low.

DR. FRANK: Was there any constant direction of change in the interview? It seems to me the most obvious assumption would be simply that the interviewee was gradually learning what kind of behavior was less stressful when the inter-

viewer maintained a standard pattern. Did he get so he got more and more comfortable in this situation?

DR. MATARAZZO: I should answer the question in this way, Dr. Frank. He changed on some of the variables and not on the others. The one he changed on was this. His units of action became longer as the number of interviews progressed. He went from very short communication units to longer units (a fact which we know happens very often in psychotherapy). There were two dimensions in which he did not change. When I failed to respond to him twelve times in the silence portion of Period 2, he took the initiative two times out of twelve in the first interview, and over the next nine interviews vascillated between two and three. In other words, his initiative was extremely low under these silence conditions at the beginning. It was still very low at the end. Likewise, to use colloquial language that is bad, in the Period 4 interruptions where I talked along with him twelve times, he submitted or stopped talking twelve times out of twelve at the beginning of the interview, and twelve times out of twelve at the end of the ten sessions.

Other dimensions, like his average duration of actions and silences, were in fact changing during the same ten sessions.

DR. FRANK: You say his utterances increased in duration. This would be a sign of getting more comfortable.

DR. MATARAZZO: We have independent evidence of one meaning of long utterances in this sense. We have studied five groups of individuals (normals and patients) who vary on what might be considered a continuum of health or adjustment. These have varied from normals to mainly neurotic out-patients, to a mixed group of in-patients and out-patients, to a group of hospitalized chronic schizophrenics. As you go from one group to another, despite huge in-

dividual differences *within* all of the groups, the *average* duration of interview actions go from very short in the schizophrenic group, i.e., the sickest in a sense, to very long in the healthiest group, the normals.

In addition to these statistically significant group differences, the interesting thing is that in any one of these groups, either the schizophrenic group or the normal group, there are individuals who are indistinguishable from the other groups. The standard deviations are quite large. When you look at *individuals* in the schizophrenic group and ask nurses and attendants on the wards to differentiate the schizophrenics one from the other, and tell you that this individual is clinically not as sick as these other patients, purely on the basis of actual *ward behavior* and independent of any knowledge of each subject's interview behavior, it turns out that the "clinically healthier" schizophrenics have interview interaction patterns which are most like the interview patterns of our normal group.

This is one of the important values of a method which is standardized and which can allow you to make these independent comparisons.

DR. ISAACS: I wanted to speak on this point of standardization also. It seems to me that the varieties of standardization of interview which are imposed on the patient and to which he responds show how readily compliant each patient is in attempting to get well or get over his discomfort. It seems to me there are many different ways of standardizing. I am not sure for myself that I can feel as comfortable with the resolution that chronological time is the same as subjective psychological time for the patient; and to make the next step, which is the important one, to the psychological meaning for the patient. I can contrast this briefly with a thing

Rosenberg, Wittert and I have been trying to do in terms of a standardized interview in a diagnostic sense.

We standardize interviews on the basis of affectively loaded components of each patient's content. We use this content for restimulation. This is, I think, according to the viewpoint of some of you a loose type of interview. To us it has very strict principles and is very closely standardized. It is reliable in the sense that each of us would do the same thing in the same situation with that patient, but each patient would be different from other patients.

Another thing that we have noticed is that the meaning of silence to the patient varies according to the context in which it is carried. No two silences are the same even for the same patient-therapist combination. At one moment a silence, such as introduced in the standard interview of Saslow-Matarazzo, may mean one thing and a few minutes later, in the next phase, may mean something quite different.

DR. LORR: Our laboratory collected initial interview tapes. In the process of analyzing the tapes one member of our laboratory computed, on a small sample, a series of correlations between the duration of speech in the initial interview, word fluency test scores and vocabulary test scores. These were surprisingly high—in the neighborhood of the 60's. This suggests that if Saslow would use a series of perceptual tests which are centrally determined, he might find some extremely interesting correlations with his interview variables.

DR. MATARAZZO: The May 1958 issue of the *Journal of Abnormal Psychology* will soon be coming out. In that you will find a paper by our group on just this point. We know that individuals show huge individual differences in their interview interaction patterns. We have asked ourselves will these individuals

show individual differences in a number of other dimensions? We have studied them by intelligence tests, the Wechsler-Bellevue intelligence test with all of its subscales, the Rorschach test, the various anxiety measures, psychosomatic measures, measures of dominance and submission, and so on, and we have found a number of very interesting Rorschach, intelligence, and other correlates of these interview interaction behaviors

DR. CARTWRIGHT: I recall somewhere in the paper, though I can't find it now, a reference to planned behavior therapy, which seems to contra-indicate what you were recently saying. To me it was a very tantalizing reference. I would like to ask you to say something further about this notion if you would like to.

DR. MATARAZZO: I would be very happy to respond to this. We have to date not used this as a method of therapy, but rather as an instrument for assessing or measuring changes which might result from a variety of interpolated activities, psychotherapy being one of these. However, there are a number of implications which have come from our own studies, and from the studies of Verplanck, Sidowski, Salzinger and Pisani, J. McV. Hunt and a number of other individuals. All of these people have shown that the behaviors of the interviewer have a profound effect on the behavior of the interviewee.

From some work that was done by Goldman-Eisler in England, and the work that Dr. Greenblatt reported this morning from his own laboratory, it is now clear to us that an individual's interaction pattern may be modified by what the interviewer does. We have reasoned in this way. Leaving our work aside for a minute, it would be possible to take off in one direction from this work and ask, "Can individuals who have very short units of interaction be made to

have longer units of interaction simply by the use of Verplanck operant-conditioning methods, or other kinds of methods?" That is, independent of content, which you would use as you would the content in any psychotherapeutic interview, just use it as it comes up, can you by increasing the duration of your own utterances, or by saying, "Uh-huh," or giving some other reinforcement to long utterances of these patients when they occur, get an individual who is at first one who speaks in short utterances to speak in longer utterances? Likewise, for an individual who has a very low initiative rate, can you by reinforcing by, "Uh-huh," every time he shows some initiative during the interview, get him, after a number of interviews, to be an individual with more initiative? The test of the validity of whatever changes you can produce in this way is what the individual does in his everyday behavior independent of the interview.

We have already shown in a study that we have done that the units that we measure in the interview are themselves highly correlated with similar kinds of behavior in real social situations. If this is true, and we can now modify what an individual does in an interview, we hope purely by the process of stimulus generation and response generalization that some of this newly acquired behavior may generalize to his extra-interview, or social behavior.

DR. SHAKOW: I could not help but react to this last suggestion by thinking back a good many years to the situation at Worcester when I think we were physiologically very naive. There was a time when we had findings which indicated that the basal metabolism rate of schizophrenics was low, lower than the normal. The obvious thing to do then was to give the patients something which would increase their BMR. Dinitrophenol had at that time become available and it

raised BMR. We gave this to the patients. But it turned out that they remained just as schizophrenic as they had been before.

Some researchers talk about increasing the affective responses of schizophrenics, the implication being that this is the way you make them well. What do they mean by that? They say, "Uh-huh," every time a patient mentions a word which has an affective significance or is related to a family response. That you can train schizophrenics to do things we know. We have had evidence of this kind for years. But the question is, is this *really* what we are talking about, or are we just talking at a surface level which is quite different? I think we ought to say with regard to the work that has been reported that it is at one level of discourse. Dr. Saslow was quite right in pointing out in reply to Dr. Greenblatt's comments, that he was not suggesting theirs as a therapeutic method.

Now, I am getting a little bit worried by the implications that have been drawn. At first the indication seemed to be that we wanted to find out experimentally what happens in a didactic situation of two people placed together in an interview situation of a kind. In such a situation what happens when you control the factors involved? Can you get certain kinds of results? I think you will come out with very interesting findings which ought to be taken into account as being probable important accessory factors in any kind of situation where two people are involved, that is in a therapy situation, a social situation or whatever else it might be. Undoubtedly one person does affect the other. It is when broader implications are drawn from such a situation that I begin to get concerned about the over-simplification which is involved.

DR. ROBBINS: There is something that has been concerning me even before I came to the conference and is very

sharpened for me in the presentations of today. I think what we have been hearing from all of our papers, yesterday and today, is something to the effect that we recognize a commonplace but we are trying to recognize it more studiously. That is, the situation in which two people find themselves and the way they behave in that situation significantly affect each other.

This has certainly been a preoccupation in the multivariable experience of let us say the analytic situation. There have been many, many papers emphasizing for instance the analytical, the psychoanalytical situation, in respect to the ways in which that induces certain reactions of the patient to serve the purposes of analysis.

I think our job is this. I like very much the tremendous range of interview types, from interviews with no therapeutic goal whatsoever that you have been describing, George, and interviews of a certain type with this therapeutic idea or goal, and other interviews of others.

You are pointing out in a very discreet and beautiful way that you can standardize a situation for yourself and you get nonstandard responses that are standard only for the individual, but not from individual to individual if you limit your situation. But then there is tremendous fluctuation, variation, and so on. We have got to be able in all of our discussions, I think, to describe and specify and understand how each one of these experiments or studies that we are making is a different, fundamentally and tremendously significantly different situation for the subjects in every instance. Because otherwise we are in the situation of having absolutely noncomparability of our experiments and our studies.

Some of the value of this is to take something out of total context. But the minute you say we are going to try to

do this with therapy, then the findings regarding standardized interviews may have little to do with the therapeutic interaction itself. The subject becomes the patient. He has an expectation. You have an expectation. The question he had, "What the hell is this all about?" that he might have had in his mind up to this point, now becomes, "This might be helpful to me," and this is a totally different situation.

We are faced with a very important problem throughout the conference, one sharpened to me today by the fact that we all have to be able to try to define the structure of the situation, its implicit and explicit meanings to the interviewer and to the subject, and to understand his responses and not try to make the error of comparing a totally different situation which has a fragment of another situation in it, or we will get no comparability whatsoever.

I hope we can keep this idea fairly clearly or we will end up with what I said to some of the others when the conference started. We will have a cellar of Babel here rather than even a tower before we are through.

DR. GREENBLATT: I could not agree more with Dr. Robbins. The situation that we have set up purports to include the interview as a didactic social interaction problem. We are concerned with the psychological dimension and with the physiological dimension.

Some of our critics felt that we ought to have our heads examined for trying to introduce so many variables in a complex field. However, they usually end up by suggesting we add one more variable—the one they are especially interested in.

DR. LUBORSKY: I would like to hear comments from Dr. Lacey or comments from the group in response to Dr. Lacey's point about physiological research along with psychotherapy. His

interest has been in understanding physiology through the information provided by the interview. I was wondering about our reaction to that. Do we expect physiological studies of this kind also to contribute to our understanding, or our technique in psychotherapy? I was thinking of one example of this, though it may not be a typical example. A patient I treated years ago would report in the course of his free associating a contraction and a sensation of secretion in the stomach. A year before he had suffered from peptic ulcers and had remained marginally free of them since. It was possible to identify in the free associations the constellation of conflicts which were associated with his report of physiological sensations.

With this patient it was possible more readily in psychotherapy to work with a major symptom, the ulcer condition, because of the information about the physiological changes. This may be a slightly unusual example in terms of the clarity of the physiological and psychological association. One could have cured the ulcer condition without having the report of the physiological state as part of the context of the free associations; one could probably have identified the conflicts in the free association material without it. It just happened this was a dramatic example in which the understanding of the patient, and therefore what was done in the treatment, was facilitated.

I would like to hear other reactions of people in this group to this question of the possible uses of physiological researches.

DR. ROGERS: Some of us in the Department of Psychiatry have been getting our feet wet in this very complex problem of physiological concomitants of psychotherapy. I thought I would just mention one incident in a hypothesis-generating attempt. In order to learn

some of the complexities I did some pilot recording of physiological variables with a client I have been working with, a professional person who is willing to be measured in that way, and who had some interest in research. Then it occurred to me after we had done a few interviews that it might be definitely helpful to get *his* reaction to the "psychological" and physiological materials.

I found that extremely informative. What we did was to play the recording with the polygraph record going by before us and I asked him to make notes of subjective reactions that he had as he saw and heard this material for the first time. I was doing the same from my point of view. I feel I learned a great deal, both about therapy and some of the possible hypotheses for further research. But to mention one minute thing that interested me. It is not uncommon for me to be rather hesitant in my speech, especially in replying to difficult material. Several times when this would happen I would start to respond, "Well, if I get that . . ." and then perhaps hesitate a moment before I was able to formulate my response. Not infrequently you would get a sharp PGR rise at the point of my silence, and his subjective reaction was quite clear. He said, "I was sure that what was coming out next would be judgmental or harsh or rejecting," and that in the face of the fact that harsh judgment is not a very common tendency on my part. It was interesting to see how readily threat could be generated by a very small silence on the part of the therapist, which I must say I would never have thought of as being threatening. But the moment it is mentioned you can readily see how it would be so. That is one tiny example of the way in which such material might feed back into the therapist's experience.

DR. SASLOW: We have some of the client-centered therapy records that you

have made available. As one listened to them with the special background of experiments that I have described, the thing which struck us was the way in which Carl Rogers would hesitate while the other person remained absolutely silent. From the viewpoint of interaction chronograph variables and swings, this meant to us he was in complete control of the interview situation, and we intend to make chronograph observations, using these variables, of a number of such published materials. There are some Yale records. There are the Rogers records. It is very interesting that the measurements based upon the interaction chronograph approach—far simpler and cruder than the kinds of statements used by Rogers to describe client-centered therapy, permit one to make certain kinds of statements which are of a totally different and at times illuminating order from the interpretive statements which come from you and the patient.

Such different approaches apparently illuminate very different facets of the total process. I think Dr. Rogers' example has brought out again that the time dimension in psychotherapy needs more attention than it has been given. How and where is all open for study.

DR. DITTMANN: I would like to talk to Dr. Luborsky's questions of Dr. Lacey and Dr. Lacey's comment that he was not sure that he could offer anything to people in psychotherapy research, but that he thought they could offer him something in the line of development of tools.

I am reminded of the experience we have had in working with professional linguists where we hoped that we would be able to get ready-made tools from them to study the ebb and flow of affect in interviews. It turns out that the linguists have developed lots of techniques for studying linguistics, but not for studying affects and they want to use us to

enlarge upon their own tools for studying linguistics in the hope that it will eventually contribute something to us. It would not surprise me at all if the physiological techniques as they exist today won't have any immediate tool value or will have value only in development of these tools until they can be of use to us. So this process that Dr. Lacey talks about, that he feels he can get more from us than we can get from him is probably a two-way business. In the course of this development we will eventually be able to get something from him

DR. WALLERSTEIN: When Dr Robbins spoke before about the cellar of Babel as something we might be getting into, and Dr. Greenblatt expressed a kind of agreement with it as if it were the consensus agreement of the whole meeting, as I thought about it, it made me think, that in a way it is shifting one of the fundamental premises of the meeting and one of its crucial problems. This goes back to what came up in the very first paper. The opening paper stated I thought one of the premises on which such a meeting is called together. I go back to Dr. Frank's paper and what he said about Eysenck. All Eysenck has been trying to tell us is that no matter what kind of psychotherapy you use, who does it, what you call it, how long it takes, or whether you don't do it at all, that everybody gets the same percentage of cures, good results and whatever, so there must be something in common that they all have until you can prove otherwise. Then a meeting like this is called together at which there are 30 people representing almost as many, if not completely different points of view, different applications of a wide variety of points of view, in part to see what they have in common. Maybe we are all blind men who are grabbing different parts of the elephant, and therefore the commonality we are looking for is fast disappearing

DR. SEEMAN: I think one of the problems, in response to Dr. Wallerstein's comment, one of the ways in which we may get congruence is to start thinking in terms more of psychological variables. I want to distinguish psychological from sociological variables. By sociological variables I mean names like psychoanalysis or client-centered therapy and others which I think have a history, but I am not sure that they deal only with behavioral variations. I think this is a thing we need to attend to.

DR. BORDIN: I have been holding back because I felt it would be greedy of me to start talking now when I get a chance to talk this afternoon. What Dr. Seeman just said is a very important point from my way of looking at it. Our problem is to identify what attributes of this process that we are studying offers a basis for choosing between theories, instead of simply labelling a process as psychoanalysis or client-centered therapy or any other name of a theory. When you start to think from any one of these frames of reference, what attributes of relationships do they make assertions about? It is important to pick out some attribute on which theories differ. I am particularly interested in mentioning one of our concepts that I think is related to both of these papers this morning. This is the attribute that arises from treating the therapist as a perceptual field. In a sense this is now taking a slightly different tack than Saslow and Matarazzo, but in a sense at the same time having a great deal of parallelism. Instead of picking out tempo patterns in the patient in response to certain uniformities, we pick out the uniformities and differences in the perceptual field that is created by the behavior of the therapist, by his office, by his couch, by his pictures and so on. This is not a specifically content defined variable. This is one of the similarities. It has to do

with whether this perceptual field offers ready formation into a clear gestalt, or whether it resists clear formulation. Here we are not thinking so much of contradictory inputs of cues as might be illustrated in a reversible figure phenomenon but of stimulus deprivation. You don't put in enough cues to offer opportunity for a clearly demanded percept. First of all we have standardized the measurement of this attribute, ambiguity. We set up a situation where we have interviewers, therapists, who are acting according to certain specifications. It is not as completely standardized as yours (referring to Saslow-Matarazzo), but we can get an almost complete non-overlapping distribution of behavior in these terms. We can then study certain reactions of the other person to this situation. In this case we were interested in arousal. This is the connection with Dr. Lacey's paper. We were interested in arousal in two senses. One, the Palmar conductance level.

DR. LACEY: You were interested in the Palmar conductance level. Why do you say arousal?

DR. BORDIN: All right, you are going to trap me on that. We were thinking about the general notion of anxiety. We were thinking in theoretical terms, partly applying the psychoanalytical formulation of the inner turmoil within an individual when he is worrying about his propensities for reaction, and when the environment does not give him the customary supports of being able to control the propensities of reaction he is afraid of. We say under these circumstances he ought to get worried. Palmar conductance is supposed to be one way that it ought to be reflected. The other kind of measure was speech disturbance.

One of the reasons why I am interested in bringing this up is that we get quite a different pattern of reaction ac-

cording to these two kinds of indices of disturbance arousal.

I was interested in how you would deal with it. The Palmar conductance level corresponds to the perceptual response of the person who is exposed to this stimulus situation. If he perceives the situation as ambiguous, in other words, somehow he does not project upon it a very clear gestalt, then this will correspond to the degree of disturbance indicated by Palmar conductance level during that interview.

The objective stimulus conditions in terms of to what extent it was a clear gestalt, corresponded now with the indices of speech disturbance. The less possibility there was of a clear gestalt objectivity defined by the situation, the more he was going to show speech disturbance.

DR. SHAKOW: I would also like to ask John Lacey why you talk so much about physiological *concomitants*. I am somewhat concerned about the use of this term here in the present context. Obviously many people are interested in the *concomitants* of the psychotherapy process. They also in some way are helping to find the answer to questions in the physiological realm. I think that is what you are criticizing and very legitimately. I think we miss a tremendous amount in not studying the physiological events which go on at the same time. But I am not using *concomitants*, that is, events which go on at the same time as the rest of the process, because I think the physiological is an important communication system. I hinted at that yesterday when I said there were four levels of communication. I am quite serious about the physiological as being an important communication system. I don't think we have any notion of how much of our reactions are due to the "awareness" at levels of unawareness of what goes on physiologically.

Mandler's work has been quite interesting in this respect—how he has been able to get people to become aware of their heart rate and other physiological responses. But whatever the case, some patients especially, I am thinking of the patient that Luborsky was mentioning, are constantly reacting to these events as part of their intrapersonal communication system. The therapist may not be aware of this, but it plays a marked role in what takes place in the therapy process.

I would like to see more emphasis placed on this area as a communication system, a communication system that is intrapersonal largely. It would be interesting to have the data about this area which we get through the kinds of experiments which you have described and which Greenblatt has described. To appreciate physiology in its non-concomitant aspect is all I am asking for

DR. LACEY: I won't take these comments in order. I will start with the most arousing comment. What Dr. Bordin said illustrated how a good investigator can be misled by a bad concept. You started out with arousal, which I think is the most elusive concept in psychophysiology today, and all too general a concept, and you ended up with something different.

If I understood you correctly, you said you obtained PGR's when there was evidence that the subject was perceiving the environment as ambiguous. Now I would suggest that you should take that one step further and ask what was happening physiologically. I would hypothesize—I don't know that this is true, we have just stumbled on this notion of facilitatory and inhibitory feedback from the autonomic effectors to the central nervous system in the last few months—I would suggest that what you were seeing there was a part of facilitatory feedback to the CNS. You were seeing

an attempt of the organism to "take in" the environment. Insofar as the results and theory at the end of my own paper at this conference are true, I would expect that you might also have observed a cardiac deceleration.

As soon as you start thinking in this way, you stop thinking of an *index* of sympathetic nervous activity. I think any critical evaluation of the literature would show that you cannot do this.

I don't think you have learned anything about the problem of anxiety or arousal. I think your bit of evidence, and many other bits, are going to teach psychophysiolgists something about sequences and patterning of physiologic activity which determine, modulate, govern and direct behavior.

I must admit I don't quite know why I reject out of hand the lie detector approach to psychophysiology and therapy. As I said in the paper, I don't know whether I am attacking a straw man or not, but I do reject the approach which states that physiology can measure a concept. A person whom I respect very greatly said to me the other day, "I think it might be useful to use PGR's as a measure of," and what he then named was a concept. How can physiology measure a concept? If such a person is challenged, he finds it very easy to slip over and say that what he is looking for is some physiological variable that varies systematically with the psychological construct that he measures in other ways, and yet what he is honest-to-God doing is using physiology to measure a concept, because he does not typically measure "drive" let us say, in other ways. He simply asserts that the autonomic variable does in fact measure motivation. What is so devastating about this is that in the history of psychophysiology you can show you can "measure" any old concept with psychophysiology. This "sea of response," this "penumbra of

activity," as Davis has called it, this "underlying fugue of response" (Kubie)—everybody is moved to poetry in talking about the autonomic nervous system—is the most common response of the organism. It is difficult to do anything with the organism without seeing an autonomic response. By the "Bavelas law" you can make of these autonomic responses anything you want to. It is like factor analysis. If you put in expectancy, if you put in ambiguity and affect, you will come out with a relationship, and thereby you miss everything that there really is to be said about the psychological importance of autonomic nervous system phenomena.

Dr. Shakow brings up another possibility, and this is a more subtle possibility. It gets away from indicant functions, lie detection. He says physiological phenomena are a source of intrapersonal communication. I am going to add that physiological phenomena may be a source of the interpersonal communication.

Some physiological events are of psychological importance. I think many of them are.

I think most of the intrapersonal communication is by reflection, visceral afference that never can emerge into verbalizable consciousness. Some of these visceral processes can be perceived. As Mandler has shown, there is a very wide range of individual differences in our sensitivity to our own guts. I think Dr. Rogers felt with his viscera and bones. I think he ought to go over and let Dr. Mandler study him, because there are people who are sensitive to their viscera and bones, and there are people who are not.

Here we have a very interesting situation where some people have a perceptual field that is more enlarged than others. This may be one of the factors—I don't know whether it is or not;

nobody knows—underlying some of the individual differences in rigidity and fluency. So this is certainly a very, very important thing. Again, this is not indicant function, or lie-detector procedure.

There is another aspect in which physiology is communication. I think there is some evidence to suggest that certain people do use physiological manifestations to communicate. There was during the war—I can't put my finger on the reference just now—a study using Rorschach's introversion-extroversion in observing what happened to aviation cadets in an altitude chamber. The extroverts, as I remember, became cyanotic and fainted, and displayed in very apparent ways that they were under distress. The Rorschach introversives did not become cyanotic, did not faint, but displayed much physiological upheaval in covert ways. The flush, the blush, the extensor activity, and the drying up of the oral mucosa—they may be ways in which the organisms announce to the world what is wrong with them or that they are feeling distress.

I say this may be one of the things that he is externalizing. Let me go on on this matter of choice. This is something nobody understands and we don't have a proper terminology for it. If you look over the papers that I have summarized at the back, and a vast number of other papers, you find that individuals have preferred modalities of response. Some will always respond maximally in a vasoconstrictor measure, a digital measure, others in terms of heart rate, others in terms of blood flow, others in terms of sweating activity, others in terms of gastric activity.

The question is "How does this all happen?" The easy way out is simply to say this is constitutional. But we don't really know this yet. This may very well be linked up to a mode of reacting of the individual, which perhaps is related

to a continuum varying from, "I announce to the world how I feel," to, "I keep it to myself." That is all I meant by that.

Dr. Luborsky brings up a most important point. If we are dealing with an ulcer patient, a headache patient, a backache patient, obviously physiological recording becomes a most relevant part and parcel of the therapeutic process. If I have a hypertensive patient and I want to know whether client-oriented therapy is effective in combating hypertension, I take blood pressure. If I have a headache-prone patient, I take neck potentials. The organization is medical here is a symptom, and here is part of the physiological mechanism related to the symptom.

Dr Dittmann says that this may be a two-way field. I think so, too. I am not really as completely pessimistic as I like to make out. I do doubt that in the current state of our knowledge psychophysiology can make a real contribution to psychotherapy. If we abandon these concepts of arousal, activation, affect, and so on, I think (maybe only because we don't know anything about these other things yet) that there is a real opportunity here. I think the psychotherapeutic situation can teach us more about psychophysiology than the reverse.

Therapy is a learning process. Those of us who like to look at therapy as a learning process can find plenty of material which we can describe in learning terms. But now learning does not go on without the participation of neural and

humoral activities. We really know nothing about them. We talk about cell assemblies or changes in synaptic resistance and a lot of other stuff. But now evidence is accumulating that the autonomic has something to do with the rapidity of learning.

Calloway in an extraordinarily interesting series of experiments is injecting amphetamine and atropine into human subjects, and observing effects on behavior. He bases his experiments on the hypothesis that these drugs change some "central sympathetic state" resulting in "narrowed attention." Now, these drugs produce a wide variety of autonomic changes, among which are heart rate and blood pressure, which I think are two of the most significant autonomic changes, because they are so clearly connected, via carotic sinus mechanisms, with the activity of the central nervous system. Calloway shows, among other things, diminution of galvanic skin reactivity to symbolic and sensory inputs—"keeping out of the environment"—as a result of methamphetamine injections. Of most interest to me is his demonstration that, while under the influence of methamphetamine, individuals cannot learn as well in a probability learning or guessing situation. We all agree that therapy is a kind of probability learning situation. It is a probability learning situation.

Now, how about this. How about certain kinds of autonomic states facilitating or retarding the learning process of therapy?

Inside The Therapeutic Hour

EDWARD S. BORDIN, PH.D.

The key to the influence of psychotherapy on the patient¹ is in his relationship with the therapist. Wherever psychotherapy is accepted as a significant enterprise, this statement is so widely subscribed to as to become trite. Virtually all efforts to theorize about psychotherapy are intended to describe and explain what attributes of the interactions between the therapist and the patient will account for whatever behavior change results. To really understand psychotherapy we must investigate in depth this more or less carefully insulated universe of behavior, geographically bounded by the four walls of the therapist's office, and temporally bounded by the patient's entry into that office, and the therapist's "our time is about up today," approximately an hour later. Conceptions of how these two individuals became the kinds of persons they are when they meet in the therapist's office must be articulated with conceptions of what can and should happen during the few or, more frequently, many hours during which they have commerce with each other and how these therapeutic events influence the kinds of persons they become after therapy. Increasingly, we are turning from an exclusive concentration on the patient and are giving attention to what the therapist brings to therapy and its possible impact on him. In the therapy project of the University of

Michigan,² we have centered our attention upon the events of the therapeutic hour. Theories of psychotherapy and the folklore of psychotherapists contain implicit and explicit statements of principles and predictions about sequences of events in therapy as a function of the behavior of either or both protagonists. Our aim is to ferret out, to clarify and finally to test the validity of these statements of principle and the predictions derived from them.

GUIDING FRAMEWORK

Before proceeding to share with you the details of our work, its fruits, its frustrations, and our current preoccupations, I think the purpose of this conference will be better served if I make explicit the major presuppositions that guide us. Too often two research workers, like patient and therapist, come no closer than hailing communication because they move in different worlds. The questions they ask, the observations they record, are so unrelated that they have little of meaning to say to each other. Therefore, I want to begin by describing what guides us in selecting the specific qualities of therapeutic interactions that we study. Also I think it is possible that these same considerations can provide pivots for an examination of alternative strategies for process research

1. Throughout this paper this term will be used to designate the recipient of psychotherapy. "Client" is an equally applicable term, especially where help is offered in a non-medical setting.

2. Supported by Grant M-516 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. Currently Dr. Richard L. Cutler and I are principal investigators. Drs. Allen T. Dittmann and Harold L. Raush earlier shared responsibility for the direction of the project.

I shall discuss three major considerations influencing our choice of variables: (a) Personality theory, (b) The interactional implications of the variables, (c) The opportunity the variable affords for testing alternative explanations of phenomena.

Personality Theory. I believe it is no accident that the first large sustained research program on the process of psychotherapy arose from Rogers' very specific and distinct theoretical orientation. Whether we deal with observations of "live interviews," electrical recordings or typescripts, the data before us is infinite in quantity. Blind explorers will either lose themselves in a morass of detail or soon become so baffled that they will turn to less frustrating areas of research. This complexity has forced us to set great store by the selectivity that theory permits. As is evident from our research, we are most heavily influenced by psychoanalytic theory. We strive to skirt the twin dangers of sectarianism and "outer directed" eclecticism. The aspects of the therapeutic hour we choose to analyze are rooted in theories of how behavior can be modified through human relationships. As Freud and Rogers, among others, have demonstrated, theories of psychotherapy cannot be divorced from theories of personality development. We are encouraged to pursue further those attributes of therapeutic relationships which have either direct translations to or indirect implications for the development of personality. Our ideal is to be able to build a network of relationships which can be traced not only through the therapist's office, but also through the laboratory, the home and school.

Perhaps the best illustration of an intimate relationship between attributes of therapeutic relationships and theories of personality development is to be found

in the collective concept of warmth with which we are currently preoccupied. More support for the presumed effect of warmth is to be found in the folklore of psychotherapists than in their formal statements. To our ultimate frustration, we started with the assumption that warmth was a unitary variable which had a profound influence on personality change. Our data soon taught us that we were thinking too superficially. The first sign that something was wrong was provided by the too slender thread of agreement we found among independent observers. When we found such ratings did not discriminate among therapists, and that the opportunity to listen to a recording of the therapeutic hour yielded warmth ratings no different from those obtained from the bare script, we were fully convinced that we must retreat to re-examine our concept of this variable.

This re-examination, which occupied over a year, and included explorations of the language of observers, and the re-examination of theoretical concepts, resulted in the selection of three distinct features of the therapist's relationship to his patient which make up part of the host of referents of the word "warmth." Since an extended statement of these new formulations has already been published (18), it would be unnecessarily repetitious to give more than a brief summary at this time. The first of these three attributes of the perceived warmth of the therapist is the degree to which he seems willing and devoted to lend himself and his resources as substitutes for those which the patient lacks or is momentarily unable to use. This we are calling the therapist's degree of commitment. The second one is the therapist's effort to understand how the patient experiences himself and the world around him. The final attribute we try to capture with the term "spontaneity," which may turn out to be as vague and diffuse

in meaning as the term "warmth." We intend it to refer to the degree to which the therapist appears free of reservations which enforce checks on his communications and his feelings as he goes about relating to the patient.

All three of these qualities are significant parts of the relationship between parents and children, as well as between therapist and patient. The helplessness of the newborn child, and the saga of his slow passage from almost total reliance on his parents toward the ideal state of mutual reliance, which should characterize maturity, has been acknowledged by almost all theorists, and has received significant attention in their formulations. The fears of abandonment, the longing to escape from challenges that seem too great to be met, the determination to recapture the infantile illusion of omnipotence, any of this excess baggage of infancy and childhood may be a part of the patient's current behavioral repertoire, and can be expected to influence his reaction to the degree of commitment offered by the therapist. It seems self-evident that the feelings of helplessness about harmonizing his inner and outer worlds, which drive the patient to seek out the therapist, will mobilize whatever anxieties he has in this regard.

Rogers has given so much attention to the effort to get inside the patient's frame of reference that it requires little further discussion. However, two points must be emphasized: one having to do with the influence of parents' struggle to understand the child and its development, and the other having to do with the definition of the attribute itself. The disparity between the infant's and the adult's commerce with the world is so palpable that there is an easy recognition that the infant's attainment of psychological individuation must rest to some degree upon his parents' willingness to treat his experience of the world as one

reality. Less frequently noted is the fact that an unbridled effort to understand becomes intrusive to the point of robbing the child of any world he can call his own. To me it is self-evident that the therapist-patient relationship is likely to re-install more vividly than most other relationships, the circumstances surrounding the parent's effort to understand the child. My second point is that there is a danger that effort to understand will become confused with accuracy of understanding, which passes under the name of empathy. To us, an effort to understand means just that: the faithfulness with which the therapist strives to gain a cognizance of the patient's views and feelings.

Our shared experience tells us that discrepancies may and do occur between what we say and what we do, or between our verbalizations of feelings and the actual expression of the feelings themselves. There is reason to believe, and a certain amount of data to support it, that these discrepancies in their parents may have an effect, sometimes profound, upon the emotional and even physiological development of children (11, 19, 22, 23, 24). Within the sphere of psychotherapy, this factor of spontaneity is recognized in the notion that a therapist cannot simply go through the motions of being interested and helpful. He must feel it. It is not just what he says, but how he says it that is important in the therapeutic relationship.

Interactional Implications. A second important consideration in choosing variables is their interactional implications. To further the goal of understanding therapist-patient relationships, we must use those variables that are laden with interactional meanings, meanings about others behavior that evokes the behavior to which the variable refers and, in turn, about the impact this behavior has on

others. Most of the variables we have chosen to study have met this criterion to some degree but it seems to apply particularly to characteristics of the therapist. In the case of one of these (2), we treat the therapist as a person, his office setting, and the task he sets the patient, as a stimulus field to which the patient must react. The therapist may create a stimulus field which offers the patient a very clear-cut configuration, or he may be very sparing in the cues he emits, thus creating an ambiguous field in a process somewhat akin to stimulus deprivation. This characteristic we have called "ambiguity." That this attribute of the therapist and his surroundings is laden with interactional implications is illustrated by research on intolerance for ambiguity and on stimulus deprivation. It is also a factor in the phenomena of projection in personality testing and psychopathology.

Many, if not most of our variables, meet several of our criteria. For example, patient resistance represents one of the possible reactions to ambiguity and at the same time has ties to the concepts of personality development through the assumed equivalence between resistance and defense mechanisms. Other natural pairings among variables we have chosen to study are as follows: the effect of therapist ambiguity on the patient's level of anxiety, the degree to which transference is a function of therapist ambiguity, the interrelationship between resistance and depth of interpretation, and the influence of the patient's interpersonal operations on therapist countertransference and anxiety.

Testing Grounds for Choice of Explanations. Finally, in our selection and definition of therapeutic variables, we attempt to honor our responsibility for studies which can offer some basis for a choice among several explanations of

the same phenomena. In another paper,³ I coined the term "pan-theoretical" to apply to those variables which allow for the diverse slants that competing theories give to that attribute of behavior in psychotherapy.

The variables, depth of interpretation and resistance, are good examples of instances where our definitions and measurements have been designed to identify those aspects of the behavior in question which are the objects of theoretical analysis and predictions. In the case of depth of interpretation,⁴ we have chosen a definition and a mode of differentiation which permits us not only to distinguish between interpretations of superficial as against those of moderate depth, such as would be necessary to test the differences in views between Fenichel and Rogers, but also makes possible the distinction between moderate and deep interpretations such as might be involved in differentiating the views of Fenichel and others like Berg or Rosen. Similarly, our measures of resistance have included allowance for the lapses and interruptions in the free-association process, which distinguishes the Freudian concept, and the patient's unwillingness or inability to accept interpretation, which is emphasized in Rankian and client-centered views of psychotherapy.

3. Delivered at the symposium on "Current Research in Psychotherapy," at the 1956 meetings of the American Psychological Association in Chicago, Illinois.

4. "Any behavior on the part of the therapist that is an expression of his view of the patient's emotions and motivations—either wholly or in part—is considered an interpretation. A patient has varying degrees of awareness of his emotions and motivations. Depth of interpretation is a description of the relationship between the view expressed by the therapist and the patient's awareness. The greater the disparity between the view expressed by the therapist and the patient's own awareness of these emotions and motivations the deeper the interpretation."

The self-imposed pressure to conceptualize within the matrix formed by competing theories has great potentials for the identification of new ways of looking at the therapeutic relationship. We believe that it was our responsiveness to this pressure that has led to the formulation of ambiguity and commitment as new variables to be studied in research on psychotherapy.

THE PROVING GROUND FOR MEASURES

For the most part, we have had to compromise with our natural desire to minimize observer error in our observations. As you can see from the examples I have cited, the requirements we set for ourselves in choosing variables, such as their relationships to personality and psychotherapeutic theory and their interactional implications, have tended to rule out observations of discrete and simplified behaviors. For the most part, our observations are derived from ratings, therefore, our progress will require skirting the many difficulties that lie along this path. We have felt that the attainment of an acceptable level of inter-observer agreement, while necessary, is incomplete evidence of the adequacy of our measures. Our approach has been to apply the notions of construct validity, which were developed by the Committee on Test Standards of the American Psychological Association and elaborated by Cronbach and Meehl (4). As Cronbach put it recently (3: p. 676), "A proposed test interpretation . . . is a claim that a test measures a construct, i.e., a claim that a test score can be linked to a theoretical network. This network, together with the claim, generates predictions about observations. The test interpretation is justified only if the observations come out as predicted." This also calls for empirical tests of the adequacy with which the method of measurement satis-

fies the requirements of the construct. This point of view added several steps in the analysis of our measures intervening between the simple establishment of agreement between observers and the application of the variable in an attempt to verify psychotherapeutic theory. One step has been to apply the latest methods of scale analysis to test our assumption that our raters were, in fact, generating a unidimensional attribute. This has been done with regard to both ambiguity and depth of interpretation and, in a partial way, with regard to resistance and patient anxiety. In the case of ambiguity, we found the clearest case of a unidimensional attribute (16). In the case of depth of interpretation, we found evidence of three attributes, two of which could not be clearly identified (17). The one that could be clearly identified and accounted for most of the variance corresponded to our conception of depth of interpretation. There were suggestions about the possible identification of one of the two residual dimensions. Two possibilities seem to be equally applicable to the vague indications we had. One was that the judges were reacting according to the degree of emotionality they anticipated the therapist's response would arouse in the patient; the other, that the judges were reacting to the degree of ambiguity of the intent of the therapist's response. Thus far, both of these hunches have been resistant to further verification.

In the case of patient resistance and anxiety, the scaling approach has not been used. Instead, a number of different ways of measuring the attribute have been intercorrelated in order to get some idea of whether the attribute was uni- or multi-dimensional. In the case of resistance, Speisman (21) studied the reliability and the intercorrelation of measures of patient exploration, opposition, superficiality, self-orientation, self-scrutiny, and

blocking.⁵ The reliabilities for ratings of blocking were so low as to force us to discard that variable. The intercorrelation among the others suggested that exploration, self-scrutiny, self-orientation, and superficiality were tapping essentially the same attribute and that it was independent of opposition. Thus, we are left with two sub-characteristics of resistance. One represents the patient's compliance with or avoidance of this task to observe himself and communicate to the therapist. The other represents his reaction to the therapist's interventions. In his studies of alternative measures of patient anxiety, Dibner (7) found disturbingly little relationship among global ratings by psychologists, the patient's self report of tension, GSR, and two different measures of disturbed speech. This only serves to illustrate the clouded state of the conceptualizations of anxiety. Merton Krause of our project, has undertaken to seek conceptual and empirical clarification of this attribute. Recent studies by Mahl and Dittes (8, 9, 14) are broadening the empirical bases upon which we can work.

One source of the problem of communication between research workers is differences in assumptions about the qualification that certain variables demand of the observer. Many research workers, particularly those of a psycho-

analytic bent, will accept as observers only those who have gone through formal therapeutic training, including personal therapy, and who have considerable experience as therapists. The issue of the qualifications of the observer would seem to apply especially to depth of interpretation. Here the observer is asked to judge the patient's level of awareness of those emotions and motivations which have been the object of the therapist's interpretive remarks. Surely the observer would need both professional sophistication and personal freedom to be sensitive to the feelings of others. If we cannot find differences between ratings obtained from sophisticated and unsophisticated observers, surely our confidence in the meaningfulness of depth of interpretation ratings would be shaken. We were relieved to discover that naïve observers, namely, undergraduates drawn from the first course in psychology, did not attain as much agreement among themselves as more expert raters, clinical psychologists who met the criterion of having had a minimum of 100 hours of experience as therapists (26). At the other extreme of the continuum of experience and personal therapy, we compared clinical psychologists and psychoanalysts in terms of rating depth of interpretation (6). Here we found relatively little difference between the groups. The clinical psychologists showed a slightly greater tendency to agree among themselves which may have been a function of their greater familiarity and readiness to cooperate in more rigidly structured rating tasks. I suppose whether this result undermines one's confidence in the method of rating depends upon one's point of view. We find the results disturbing because they suggest that the ceiling on how much information a clinician can integrate may be so low as to be unresponsive to differences in amount of training and experience after a certain minimum has been reached.

5. Defined as follows.

Exploration: The degree to which the patient explores thoughts or feelings.

Opposition: The degree to which the patient is oppositional toward the therapist or the therapeutic process.

Superficiality: The degree to which the patient's comment is superficial as opposed to meaningful.

Self-orientation: The degree to which the patient is concerned with himself.

Self-scrutiny: The degree to which the patient is reacting to his own verbalizations.

Blocking: The degree to which the patient gives cues that he is minimizing or unable to comment on an idea."

A study by Townsend (25) offers us an illustration of another source of contribution to construct validity, in this case with regard to ambiguity. An ambiguous stimulus field does not readily lend itself to one interpretation. Therefore, we can assume that therapists who are rated as being more ambiguous ought to be perceived with greater variability than therapists who are rated less ambiguous. Townsend, applying the Q-sort technique, asked naïve observers to sort descriptive adjectives according to the degree to which they applied to the therapist, after hearing a section of one of his interviews. The intercorrelations of Q-sorts by ten raters were distinctly lower when the therapist had been independently rated as behaving in a more ambiguous manner. Recorded excerpts of interviews by four different therapists yielded rankings in terms of average intercorrelations that corresponded exactly to their rank with regard to rated ambiguity. Thus, we are strengthened in our belief that our measures of the ambiguity of the therapist do in fact reflect the kind of attribute we had set out to measure.

A final source of construct validity, drawn from studies of the measurement process itself, is the investigation of the influence of context on ratings obtained and on agreement between observers. The application of this kind of evidence was illustrated earlier. I then discussed how we were helped to abandon the global conception of warmth by the failure to find that the opportunity to listen influenced ratings of that quality. Similarly, we have found that the total elimination of context, i.e., rating therapist's responses for depth of interpretation, without any knowledge of the intervening patient communications or of the sequential order of the responses to be rated, does influence ratings (12). This, of course, reassures us that expert raters are in fact being influenced by an opportunity to perceive something about the

patient's level of awareness. Here again we are left with doubts because excising the intervening patient communication does not by itself influence our experts' ratings. It would appear that raters are incapable of using more than a slight amount of information.

PATTERNS OF EXPERIMENTAL DESIGNS

The heart of construct validity is to be found in demonstrable relationships between any given variable, and other observations drawn either from within or outside the therapeutic process. We achieve the greatest degree of confidence in the validity of a given measure when we can verify predictions formally derived from specific assumptions about the concept in question. Our research strategy has been to give considerable attention to the internal analysis of our measures and to test assumptions about the condition of measurement before moving to a greater stress on external validity. Consequently we have accumulated relatively less data at this level of construct validity. However, we do have a few completed studies which illustrate the different kinds of patterns that might be followed. Over the last year, there have been an increasing number of studies produced at other centers which have also illustrated these types of experimental designs.

Speisman, in the study mentioned earlier (21), went on to test assumptions about the relationships between depth of interpretation and resistance. He drew two samples of interpretation-resistance sequences. One was made up of groups of eleven successive responses randomly selected from portions of 21 cases involving 21 different patient-therapist pairs. The other population represented an analysis of all responses in a group of 5 successive interviews from one case. Essentially the same results were obtained from both populations. He found clear-cut evidence that patient resistance

was lowest following interpretations at the moderate level and less conclusive evidence that superficial interpretations were followed by less resistance than followed deep interpretations. These differences, according to level of interpretation, were brought into even sharper focus when the analysis was done by shift in level as compared to absolute depth.

Dibner (7), like Speisman, relied mainly on the relationship between two variables drawn from within the therapeutic hour. He tested the prediction that ambiguous therapeutic relationships will provoke patient anxiety. The choice of a quasi-therapeutic situation enabled Dibner to introduce special experimental controls. The experiment was carried out in a general hospital utilizing a diagnostic interview which was part of the patient's admission to the psychiatric service. Four clinicians interviewed ten patients each, half of them under the instruction to be more ambiguous, and the other half under the instruction to be less ambiguous. Tape recordings provided the data to check the accuracy with which these instructions were carried out. Measures of anxiety were obtained in the form of ratings of speech disturbance, PGR conductance level, (continuously recorded during the interview) and global ratings of anxiety. In addition, a retrospective report was obtained from the patient following the interview and this provided data from outside the interview on the anxiety he had experienced and his perceptions of the interviewer. As I have mentioned earlier, there was a disturbing lack of agreement among these different measures of anxiety. As a result the pattern of verification of our predictions was not clear-cut enough to foster a high level of certainty even though there was a considerable amount of statistically significant verification of the expectations demanded by our theory. Cutler and

Rigler's work on transference illustrates interesting and ingenious experimental designs which will have application in many other settings. Cutler (5) developed a crude but significant way of identifying one aspect of potential countertransference in the form of conflicts in the therapist's modes of interpersonal relations. He did this by comparing the therapist's self-ratings of sixteen interpersonal characteristics based upon the Leary, *et al*, methods of coding interpersonal relationships with ratings obtained from the therapist's associates, (13). A conflict area was defined as one where the therapist either over-emphasized or under-emphasized his characteristics in comparison to his associates' view of him. It was in these areas that the therapist showed corresponding over- and under-sensitivity in reporting his own interactions with the patient. In this case, his report was compared with a coding of the typescript of recordings of his interviews; further, the therapist tended to be less accurate in reporting those aspects of the patient's interactions with him that were conflict laden for him as compared to those not laden with conflict. Finally, the therapist was found to be more defensive in his reactions to those patient maneuvers that impinged upon his areas of conflict.

Rigler (20) extended this design to investigate whether or not the therapist would give physiological indications of anxiety in relation to the production of conflict relevant material by the patient. Rather than risk an inappropriate intervention into the therapeutic relationship, he arranged to have two therapists listen to recordings of their own therapeutic interviews within a day or two afterward. At this time continuous PGR records were obtained. Using Cutler's methods of identifying conflict areas, he found significant relationships between these physiological indications of anxiety in the therapists and the appearance of con-

flict relevant material in the patient's production. Another feature of Rigler's study was the attempt to test the influence of a therapist's anxiety on his tendency to remain more or less ambiguous. We had thought that anxious therapists would need to increase structure as a way of trying to control the patient's anxiety provoking behavior. The failure to obtain confirming results forced us to re-examine our reasoning and to realize that our predictions were rather hasty in that they overlooked variations between therapists in methods of dealing with anxiety as well as variations within a therapist according to the type of material

As we all know, one of the biggest obstacles to research on psychotherapy is the limitation on the experimental manipulation of the situation which the therapist's devotion to his patient's welfare usually imposes upon him. We are currently very much intrigued by our conceptualization of commitment because it seems clear that this is a characteristic of therapeutic relationships which can be transferred into experimental non-therapeutic inter-personal relationships with some reasonable expectation of equivalence. Any relationship to a helper, whether to teacher, parent, friend or other, should offer an opportunity to test assumptions about the relationship between personality and reactions to help. Briefly, we are interested in two kinds of children. One is the dependent child who is fearful of having to rely on his own resources and is ready to escape from the challenge of using them. The other, we call the counter-dependent child, who also is in fear of being on his own but hides this fear by an exaggerated independence. When placed in a problem solving situation with an adult helper, our theories lead us to expect that counter-dependent children will be threatened and distracted from the problem solving task by the

helper's offers of aid. These children will divert their energies from problem solving into an attempt to keep the adult from helping them. On the other hand, we expect that the dependent child will divert his energy from problem solving into soliciting help where the adult seems disinclined to offer very much. This kind of study, now in progress, is being joined to a study being done by Townsend in which he is investigating our expectations of the relationship parental attitudes and child-rearing practices have to these aspects of the child's personality.

A brief description of one twelve-year-old boy drawn from preliminary studies may serve to illustrate the kind of data we are collecting. Jack was rated by his teachers as one of the most over-dependent children in our small preliminary sample. One of the observers described him as "a small, scared rabbit, very defensive and restless." On twelve story completion items, intended to tap his attitudes toward receiving help from others when he faces a difficult problem, he consistently (nine of twelve) offered answers which point toward a reliance on others to solve his problems. The boys in his story completions get angry when adults ask them to do difficult things but readily accept or even solicit aid from these adults or even from peers. In fact, he has great confidence in this aid; his boys when helped win sailing races and become very proficient in diving and otherwise achieve satisfying outcomes. Jack's mother talked about the importance of strictness, orderliness, and neatness. She wanted her children to be responsible in the sense of coming directly home from school and generally being where she expected them to be. Though she spoke of her own parents as having been too strict, and of herself as being too lenient, her responses to specific questions gave the impression that freedom was a dangerous thing to give children. She reported that she wants to

know where her children are and what they are doing at all times. In the problem solving situation, Jack was one of the few boys who asked the experimenter for help even before his first offer of aid. Even though he did refuse offers a few times, he usually accepted help readily. He continually asked questions designed to elicit signs of encouragement and approval.

If we can establish predictable relationships between personality measures and reactions to a uniform amount of help in problem solving, we are ready for the next step, namely to systematically vary dependency and amount of aid.

At the same time, the parallel question, how a patient's personality influences his reaction to help from the therapist, is being studied by Miss Joan Williams in the context of the patient's use of the therapist's interpretations. She is developing a measure of degree of therapist commitment. Her plan is to compare over-dependent and counter-dependent patients. She will be trying to determine how degree of therapist commitment influences the patient's reaction to the therapist's interpretation. We are inclined to expect that the counter-dependent patient will need to reject the therapist's interpretations, especially if the therapist seems to be accenting his promises to be of aid. Conversely, the dependent patient will more likely interrupt the process of self-exploration when the therapist seems to be leaving him on his own. If our expectations are fulfilled in these studies, we will have made significant progress toward confirming the network of empirical implications of our theories about commitment as an aspect of therapeutic work.

PROSPECTUS FOR THE FUTURE

It is self-evident that we, like those who investigate other aspects of psychotherapy, cannot hope to unravel rapidly the many mysteries and perplexities

which surround this highly complex enterprise. Progress can only be made through the slow painstaking work of perfecting and proving the variables we measure. As we look at our own work and that of others, we are impressed by the fact that many scientists are working on closely related variables, yet their definitions and their methods of measurement are sufficiently different as to preclude direct translation of results from one study to another. One of the major achievements of a conference of this sort should be to examine these slight but retarding differences so that we can learn whether they flow from resistances to modifying the familiar or from genuine differences in theoretical orientation. If the latter is the case, it should then be possible to identify and plan those studies which will help us decide which set of presuppositions seem most applicable to the variables to be measured and to the process of measurement. Thereby, we can move to a greater communality and interchangeability of measures. Most important of all, such action should bring alternative theories into direct contact in the arena of research.

We continue to be troubled by the negative results we have obtained in studying the influence on our raters of personal therapy and of professional experience as therapists. The answer may lie in the visual cues which we have not been able to introduce into our experimental designs. Perhaps those few centers that have facilities for either direct observation or the use of sound motion pictures will be able to give the answer. The other possibility we have not investigated is that we have not given the observer sufficient historical context. If our psychoanalysts were rating the seventy-fifth interview for depth of interpretation, after having had the opportunity to follow the patient through all of the preceding interviews, then perhaps they would have been able to bring to

their ratings considerably more than our clinical psychologists under the same conditions. Dedicated as we found both psychoanalysts and clinical psychologists to furthering the cause of psychotherapeutic research, we have not been in the position to ask them to devote the amount of time required to test this question in the way described.

Finally, it should be clear that many of the most significant aspects of the process of psychotherapy we have either left unexplored or investigated only superficially. For example, transference has eluded us thus far. Similarly, when psychotherapists talk about interpretation, they usually treat it in a much more specific way than we have in our studies. Interpretation usually takes into account what it is that's being interpreted, the particular mode of defensive operation, and the specific anxiety-laden impulses that stimulated these defensive operations. Further, the therapist, particularly the psychoanalytically oriented one, speaks of the change as therapy progresses in the form of the derivatives of his conflicts that the patient produces. As therapy progresses, the patient no longer finds it necessary to produce as distorted derivatives of his conflicts (10) and thus eventually is able to face himself and his impulses in a direct and realistic way. Following up all of this calls for intensive case study. We are currently exploring extensions of the methods used by Murray (15) as ways of coming to grips with these more subtle phenomena through the intensive analysis of an extended therapeutic relationship. We will need to accumulate a large enough series of these studies to provide sufficient replications for conclusions that can be generalized to apply to populations of patients and therapists.

We believe that when such a series of studies can be carried through, we will

have arrived at a point where greater strides toward understanding the process of psychotherapy are possible.

REFERENCES

1. Auld, F. Jr., & Mahl, G. F. A comparison of the DRQ with ratings of emotion. *J. abnorm. soc. Psychol.*, 1956, 53, 386-388.
2. Bordin, E. S. Ambiguity as a therapeutic variable. *J. consult. Psychol.*, 1955, 19, 9-15.
3. Cronbach, L. J. The two disciplines of scientific psychology. *Amer. Psychol.*, 1957, 12, 671-684.
4. Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, 52, 281-302.
5. Cutler, R. L. Countertransference effects in psychotherapy. *J. consult. Psychol.* (in press).
6. Cutler, R. L., Bordin, E. S., Williams, Joan & Rigler, D. Psychoanalysts as expert observers of the therapy process. *J. consult. Psychol.* (in press).
7. Dibner, A. S. The relationship of ambiguity and anxiety in a clinical interview. Unpublished doctoral dissertation. Univer. of Michigan, 1953.
8. Dittes, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, 54, 187-191.
9. Dittes, J. E. Galvanic skin response as a measure of patient's reaction to therapist's permissiveness. *J. abnorm. soc. Psychol.*, 1957, 55, 195-303.
10. Dittmann, A. T. & Raush, H. L. The psychoanalytic theory of conflict: structure and methodology. *Psychol. Rev.*, 1954, 61, 386-400.
11. Escalona, Sibylle K. Feeding disturbances in very young children. *Amer. J. Orthopsychiat.*, 1945, 15, 76-80.
12. Harway, N. I., Dittmann, A. T., Raush, H. L., Bordin, E. S. & Rigler, D. The measurement of depth of interpretation. *J. consult. Psychol.*, 1955, 19, 247-253.
13. Leary, T. *Interpersonal diagnosis of personality*. New York: Ronald Press, 1957.
14. Mahl, G. F. Disturbances and silences in patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, 53, 1-15.

15. Murray, E. J. A case study in a behavioral analysis of Psychotherapy. *J. abnorm. soc. Psychol.*, 1954, 49, 305-10.
16. Osburn, H. G. An investigation of the ambiguity dimension of counselor behavior. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
17. Raush, H. L., Sperber, Z., Rigler, D., Williams, Joan, Harway, N. I., Bordin, E. S., Dittmann, A. T. & Hays, W. L. A dimensional analysis of depth of interpretation. *J. consult. Psychol.*, 1956, 20, 43-48.
18. Raush, H. L. & Bordin, E. S. Warmth in personality development and in psychotherapy. *Psychiat.*, 1957, 20, 351-363.
19. Ribble, Margaret A. *The rights of infants*. New York. Columbia Univer. Press, 1943.
20. Rigler, D. Some determinants of therapist behavior. Unpublished doctor's dissertation, Univer. of Michigan, 1956.
21. Speisman, J. C. The relationship between depth of interpretation and verbal expressions of resistance in psychotherapy. Unpublished doctor's dissertation, Univer. of Michigan, 1956.
22. Spitz, R. A. Hospitalism. *Psychoanal. study of the Child*, 1945, 1, 53-74.
23. Spitz, R. A. Hospitalism: A follow-up report. *Psychoanal. study of the Child*, 1946, 2, 113-117.
24. Spitz, R. A. & Wolf, Katherine M. Anaclitic depression. *Psychoanal. study of the Child*, 1946, 2, 313-342.
25. Townsend, A. H. An empirical measure of ambiguity in the context of psychotherapy. *Mich. Acad. Science, Arts and Letters*, 1956, 41, 349-355.
26. Williams, Joan. Unpublished study in files of Therapy Project, Univer. of Michigan.

Some Investigations of Relationship in Psychotherapy

WILLIAM U. SNYDER, PH.D.

Since this meeting was organized for the exchange of views on therapy research and for reporting the research we ourselves are currently engaged in, I will not attempt to refer to all of the relevant studies which have been done by other researchers on the therapy relationship, but rather will concentrate more on the work we have been doing at Penn State.¹ A few other research centers are carrying on slightly similar kinds of investigations, but representatives of those centers are present at this meeting, and know more about their work than I do.

One way to begin an analysis of the therapy relationship is to try to state in the simplest possible terms the factors that are involved. It seems to me that some of the most important are: (a) the characteristics of the client, (b) the characteristics of the therapist, (c) what the therapist does, and (d) the interaction between these factors. One might also include other factors such as the degree of need of the client to change, or the expectations the client brings with him as to what the therapy process will involve, or the effect of social class on the relationship. However, I will discuss only the first four of these, having no data to present concerning the latter ones. We have done some studies in each of the four main areas I have just listed, and our most recent studies have dealt especially with the interaction between some of these factors.

1. All but two of the studies mentioned in this report were directed by the writer. Of the remaining two, one is being conducted by him alone, and the other by him in conjunction with three colleagues, Drs. Ford, Urban, and Ray.

As to the theoretical frame of reference for our studies, one could approach a study of the therapy relationship from any of three or more theoretical systems. Freud was strongly convinced of the significance of transference to the therapy relationship; Rogers emphasizes the effectiveness of warmth and uncritical acceptance of the client; and Dollard and Miller propose that a therapist needs to be "mentally free, empathic, restrained, and positive" and they caution that the therapist must be constantly on guard against unconscious motives of his own which may interfere with therapy. We have attempted to find a common denominator—i.e., to design studies which deal with basic aspects of the therapy relationship and which would be explainable in terms of all three of these theoretical systems.

PERSONAL CHARACTERISTICS OF THE CLIENT WHICH CONTRIBUTE TO THE RELATIONSHIP

I should like to discuss first some of our findings about the characteristics of the client's personality which contribute to the relationship between client and therapist. In our first group of studies, *Core Group I* (10) completed in 1951, there were two studies which pioneered in this area.

The group of nine studies were designed: (a) to evaluate factors relating to the predictability of success in client-centered therapy, (b) to analyze and show the interrelationships between measurable characteristics of client-centered therapy, and (c) to analyze aspects of counselor personality or ability which may relate to success in client-centered therapy.

The procedure for the studies was as follows: there were 353 phonographically recorded nondirective interviews, of which slightly more than 60 per cent were transcribed for analysis. Pre- and post-therapy tests administered to the subjects were the Rorschach, the MMPI, and the Mooney Problem Check List. Also, two post-therapy rating scales of symptoms and/or outcome of therapy were devised. Ratings were made of clients, by counselors, and by independent judges. Reliability of the ratings was demonstrated. Counselor techniques in the interviews were reliably classified. A multiple criterion of measures of success was produced.

In one of these nine studies, Gallagher (4) showed that a group of clients who dropped therapy after one or two interviews was no less maladjusted than the group of clients who continued therapeutic contact, and that both of these therapy groups were decidedly more maladjusted than comparable groups of college students who did not come to the clinic for psychotherapy.

The group of clients who continued therapeutic contact appeared to show, as predicted, more anxiety at the beginning of therapy than did the group that left therapy.²

The group of clients who left the therapeutic contact appeared to show more defensiveness, particularly as represented by measures of productivity on tests, than did the group that continued therapy.

It had been suggested earlier that the reason for clients leaving therapy could be that the nondirective counseling situation did not contain enough inherent positive rewards to counteract the nega-

tive feelings aroused within some clients. Specifically, it was considered possible that the counselors, in their strong desire to be as nondirective as possible, created an atmosphere that was a little colder or more impersonal than they would have wished.³ It must be stressed that this, *per se*, is not a fault confined to nondirective counseling, but rather that it may very well happen to beginning counselors in almost all techniques. It might be of value in the future to train students specifically in techniques designed to create an atmosphere of warmth and understanding. It would be predicted that fewer clients would be likely to leave under such circumstances.

The second relationship study in that first group was Gillespie's (5) analysis of resistance, one of the first ever conducted.

Gillespie analyzed verbal signs of therapy resistance in typescripts of 218 client-centered interviews of 43 cases. Various resistance measures were developed and related to length of treatment, success of therapy, and certain directive counselor techniques. He drew the following conclusions, among others, concerning verbal signs of resistance in client-centered therapy:

(1) Verbal signs of resistance toward the therapist and the therapeutic process, excluding within-client signs,⁴ are preceded by counselor directiveness.

3. I do not wish to imply that client-centered methods of therapy are inherently colder than others. In fact, Rogers has clearly demonstrated the opposite in his work. Also, I should point out that there have been some studies with more experienced therapists than we had in this first group.

4. Gillespie's "within-client" resistance is the classic clinging to the neurosis which constitutes the majority of non-therapeutic behaviors of most clients. The other forms of resistance were resistance to the therapist, and resistance to the method. It is recognized, of course, that the latter two may be only subtle forms of primary or "within-client" resistance.

2. Measures used by Gallagher included the Elizur Anxiety and Hostility scales on the Rorschach, a composite MMPI maladjustment scale, the Taylor Anxiety scale, and the number of items checked on the Mooney Problem Check List.

(2) Total verbal signs of resistance toward the therapist, the therapeutic process, and within the client are not preceded by counselor directive statements.

(3) Over half of the 43 cases show a decrease in resistance from the first third to the last third of treatment, but this decrease is not correlated with "success" in therapy.

(4) Verbal signs of resistance tend to be proportional to the number of counselor statements, regardless of the counselor category.

Gillespie's implication that all resistance is "bad" might be accepted by Rogers, but would not be credited by some other schools of therapy. For some analysts, resistance might be considered evidence that an interpretation was accurate but somewhat threatening to the ego.

Although Gillespie seemed concerned quite a bit with counseling technique, I have classed his study as pertaining to the client characteristics and their effect on the relationship, because he dealt with the within-client signs of resistance, and the resistance to the therapist. Resistance to technique proved to be only a minor part of the resistance evident. Gillespie demonstrated that these resistant characteristics of the client do have a marked bearing on the therapy relationship.

Much more material on the client-therapist relationship comes to us from our *Core Group II*, completed in 1956. This group of four integrated studies (1) conducted by Ashby, Ford, Guernsey, and Guernsey was designed to examine differences in the relationship formed between clients and their therapists during the process of two brief verbal psychotherapies.

The effects of two independent variables were examined in these studies. The first was the verbal behavior of the therapist. Therapists were required to administer either a leading or a reflective type of psychotherapy. The second was

the therapist as an individual, insofar as therapists had a differential effect on their clients in some manner other than through their verbal techniques.

The dependent variables included:

(1) Four measures of the relationship which was formed between client and therapist during the course of psychotherapy: a Client Personal Reaction Questionnaire containing 40 items reflecting clients' defensive reactions to therapy or therapists; 40 items reflecting clients' positive reactions; and a Therapist Personal Reaction Questionnaire of 35 items reflecting therapists' negative reactions to therapy and client; and 35 items of positive reactions.

(2) Pre-therapy measures of client's autonomy, succorance, deference, dominance, aggression, and cognitive ambiguity (based on Edwards, MMPI, Mooney, and Siegel TICA tests.)

(3) Six measures of client reactions in the therapy interviews: these were the amounts of verbally expressed dependency, guardedness, covert resistance, overt resistance, openness, and defensiveness.

(4) Variables designed to measure therapist personal characteristics: (a) ability to enter the phenomenological field of another, (b) sympathetic interest, (c) acceptance of others, (d) social stimulus value, (e) need for aggrandizing the self, and (f) aggressiveness.

Therapists were ten advanced graduate students. Clients were 40 individuals who came to The Psychological Clinic of the Pennsylvania State University between February and May of 1955. Most of the clients were young adults whose symptoms were primarily neurotic in character. Twenty clients were assigned to each treatment method.

The same therapists administered both therapies, having been given special training in the differentiation of the two. Which treatment would be administered to which client was determined by the

random assignment of clients to both treatment and therapist by means of a table of random numbers.

These four studies by Ashby, Ford, Guernsey, and Guernsey were completely integrated, and have been reported jointly; it is feasible to report results without in each case indicating the specific author responsible for them. One focus of the studies was to determine whether the pretherapy characteristics of clients relate differentially to the clients' reactions to a reflective and to a leading type of therapy.⁵ Pretherapy defensiveness and pretherapy aggressiveness did relate differentially for the two therapy techniques, leading and reflecting, to the amount of verbal behavior of clients in therapy. Similarly, pretherapy needs for deference and autonomy related to the subjective defensive reactions of clients. Clients who were more aggressive when they entered therapy tended to react more defensively in the leading therapy and less defensively in the reflective therapy; likewise, clients who entered therapy with tendencies to be deferent. On the other hand, clients who entered therapy with more need for autonomy tended to feel less defensive in leading therapy. With regard to reflective therapy, most of these relationships were not apparent.

In a current follow-up of the Core II studies, Ford is conducting a series of factor analyses to examine (a) the dis-

tinguishing characteristics of any groups of clients who react in a characteristic defensive or positive manner regardless of the type of therapist, and (b) the distinguishing characteristics of any groups of therapists who elicit a characteristic positive or defensive pattern in their clients, regardless of the types of client. This analysis will be repeated on data obtained on the same client-therapist pairs at two different points in therapy, to observe the stability of the factors. It is believed that this type of research pattern might eventually make it possible to pair clients and therapists, in order to achieve predictable consequences in the therapy relationship.

Summarizing this section of the discussion dealing with the effect of certain client variables on the therapy relationship, we have been able to demonstrate some of the client characteristics which make a difference. They are: (a) anxiety, (b) defensiveness, (c) resistance to the therapist, (d) "neurotic resistance," (e) aggressiveness, (f) need for deference, and (g) need for autonomy. We have probably shown that these client characteristics make more of a difference upon the relationship in leading types of therapy than they do in the reflective types of therapy.

PERSONAL CHARACTERISTICS OF THE THERAPIST WHICH CONTRIBUTE TO THE RELATIONSHIP

Turning to the therapist characteristics which we have been able to demonstrate to have an effect on the therapeutic relationship, I redirect your attention to one study from among our *Core Group I* (1951), that of Aronson (2).

Aronson demonstrated that, although four different counselors differed significantly among themselves on certain variables, significant differences were not observed among the clients counseled by the different counselors with regard to

5. This question was examined with regard to each of the following client pretherapy characteristics: (a) need for autonomy, (b) need for deference, (c) need for succorance, (d) need for aggression, (e) tolerance-intolerance of cognitive ambiguity, and (f) defensiveness. The relationship of each of the preceding variables to client reactions to leading and to reflective types of therapy was explored with respect to (a) the therapeutic relationship as viewed by clients, and (b) the amount of defensive verbal behavior exhibited by clients in therapeutic interviews.

age, length of treatment, estimated maladjustment, nor on five other variables, including verbal behavior, self-ratings, and outcome of therapy.

While the counselors did not differ significantly from each other on scales designed to measure understanding of self and understanding of clients, when evaluated by their peers, they were rated as differing significantly among themselves on insight into self and insight into others, and the counselors differed significantly from each other in the use of directive and nondirective techniques in therapy; the more directive counselors showed more self-insight and insight into others.

The counselors differed significantly from each other on 24 personality characteristics, which were reduced to six groupings of traits. The six personality groupings (which were correlated with the nondirective score) were adaptability, emotional control, social interaction, dependence, intellectual curiosity, and achievement. The last two groupings correlated very weakly.

Some pertinent data on significant counselor characteristics come to us from *Core Group II*. The view that individual therapists create different effects on their clients independent of the type of therapy given is partially supported by these four studies. Therapists in the *Core Group II* studies differed significantly in the defensive and positive feelings they elicited from their clients during the first eight interviews, regardless of which treatment method they were employing. They also differed significantly in the extent of guarded verbal behavior exhibited by their clients in their first four interviews. However, the view that selected therapist characteristics are related to the kinds of relationships established, the amount of defensive or guarded verbal behavior elicited in clients, and the amount of

change in client adjustment, was not supported.⁶

We must say that while this study demonstrates that therapists' differing characteristics do affect some aspects of the relationship, we are not able to say what specific characteristics of the eight studied might be the significant ones.

In 1957 four more studies, known locally as *Core Group III*, were completed. In conception and methodology they were identical with *Core Group II*, although different hypotheses were being considered. Two of these bear on the question of therapist characteristics, and their effects on the therapeutic relationship.

Karmiöl (7) was able to demonstrate only one relationship between the degree of therapists' preference for either a leading or a following type of therapy and the client's positive or defensive responses to therapy, or the therapists' positive or negative response to the client. A statistically significant and negative correlation was found between the reflective scale of the Therapists' Conception of Good Therapeutic Handling Test and the negative scale of the Therapist Personal Reaction Questionnaire. This seems to indicate that therapists who prefer a client-centered therapy do

6. The therapist personal characteristics explored were (a) ability to enter the phenomenological field of another, (b) sympathetic interest, (c) acceptance of others, (d) social stimulus value, (e) need to aggrandize the self, (f) aggression, (g) positive views of the relationship, and (h) negative views of the relationship. The client variables measured were (a) defensive verbal behavior, (b) positive attitudes toward therapist, (c) defensive attitudes toward therapist, (d) changes in maladjustment, (e) dependence, (f) openness, (g) guardedness, (h) overt resistance, (i) covert resistance, (j) anxiety, (k) positive attitudes toward self, (l) positive attitudes toward others, and (m) therapists' evaluations of client changes,

not tend to express negative feelings toward their clients in the early stages of therapy. This evidence gives some support to the idea that there is a relationship between a therapists' preference for therapy method and his response to his clients.⁷

Kahn (6) found that, insofar as therapist perception of client dynamics and behavior is concerned, the therapist who is more negatively disposed toward one client than another perceives the dynamics and behavior of the former less accurately in reflective therapy, and there is a similar trend in leading therapy. A trend toward more accurate perception of client dynamics is noticed when the therapist feels positively toward the client in leading therapy. No support is found for this idea in reflective therapy.

The passive-active qualities of a therapist are related to the accuracy of his perceptions of the client in a leading-type therapy. Preference for a type of therapy is related to the therapist's ability to judge and/or empathize with the client. Leading and reflective therapies affect differentially the accuracy of therapist perceptions of client dynamics.⁸

In the Kahn study, and to a slight extent in the Karmioli study, we are beginning to be able to label some of the therapist characteristics which make a difference in the relationship, *i.e.*, (a) negative attitudes toward the client, and (b) preference for a type of therapy in terms of passive versus active characteristics.

Peterson (9) conducted an independent study designed to determine the

nature of common patterns of preference for certain types of evidence of therapeutic gain in their clients. Thirty-five therapists trained since 1950 at the Pennsylvania State University were the subjects used. These were randomly selected from a pool of 54 therapists, constituting 96% of those trained by the University during the period named.

Fifty-four client statements were devised to represent the kinds of responses which therapists might judge to be possible indications of therapeutic gain. The composition of the client statements was controlled so as to give equal representation to eleven important characteristics within a Q-sample. The sorting task required the therapists to make judgments of the potential significance of the client statements such that if the therapist gave more attention to or followed up the client responses, greater therapeutic gain might be expected. Q-sorts which reflected judgments of therapeutic gain by the 35 therapists were factor analyzed, and six factors were obtained. The interpretation of these factors was based upon items from the structured Q-sample.

Each item of the Q-sample was correlated with each of the six factors. An inspection of the structure of the items which correlated highest with each factor yielded the most preferred and rejected characteristics. The interpretations of the six patterns of preferences and rejections indicated that the therapists selectively attended to client responses in terms of (a) concern over dependence, (b) concern with conformity, (c) concern over sex, (d) feelings of aggression but undercontrolled, (e) feelings of aggression but overcontrolled, and (f) aggression and ambivalence.

Summarizing the area of personal characteristics of therapists which affect the relationship, our findings are rather scattered. But we have been able to demonstrate that the following bear on

7. Karmioli found no relationship between therapists' acceptance of a particular therapy method and their confidence in predicting the client's behavior.

8. A somewhat similar study at Penn State by Mazurkiewicz (8) was unable to demonstrate positive correlations between type of therapy and changes in client's concept of self, of ideal, and of the therapist.

client behavior to some extent: (a) self-insight, (b) insight into others, (c) adaptability, (d) emotional control, (e) social interaction, (f) dependence, (g) negative attitudes toward the client, (h) preference for a given type of therapy (active or passive), and perhaps the following three characteristics: (a) concern with conformity, (b) concern with sex, and (c) concern with aggression.

We are clearly able to say that counselor variables do affect the therapy relationship, but we are considerably less certain as to which variables are the significant ones. (This matter will be discussed further under the section on the interaction between therapist characteristics and methods)

THE EFFECTS OF THERAPY TECHNIQUES UPON THE RELATIONSHIP

Considering the effects of the therapy techniques upon the relationship, much of what has been discussed in the earlier portions of this paper is relevant. However, in two situations very specific attention was given to this matter. In the *Core Group II* (Ashby, Ford, and the Guerneys, 1956), the question was asked whether leading and reflective type therapies produce a differential effect on the client-therapist relationship. It was only slightly supported: of twenty-one variables examined, only client guarded verbal behavior, and the therapists post-therapy attitudes toward the client, were the ones showing a significant difference between the two methods.⁹ This is perhaps not much more than could be expected on the basis of chance. It would be a very discouraging finding, were it not for the fact that, as I shall describe later, much more significant

findings were obtained when consideration was given to the interaction between therapist and his method, and the effect of this covariance was analyzed.

One study in *Core Group III*, that of Baker (3), was directed at the problem of varying the therapeutic technique. The major purpose of the experiment by Baker was the investigation of the differential effects of a leading and of a reflective psychotherapeutic approach on indiscriminate perceptions and resistance to analyzing problems.¹⁰ The statistical analyses supported the following hypothesis: that a leading psychotherapy is more effective than a reflective psychotherapy in reducing personal overgeneralizations. Inspection of the data suggested that persons who discontinue psychotherapy will reveal more resistance to analyzing problems than those who continue psychotherapy.

It was assumed, *a priori*, that implicit in a leading psychotherapy were a complex of techniques which drew the client's focus onto his misconceptions or overgeneralized concepts. If it can be assumed that overgeneralization can serve as a defense against anxiety, then perhaps the importance of maintaining such behavior can be better understood. (However, a leading psychotherapy was found more effective than a reflective psychotherapy in reducing indiscriminate perceptions, discrepancies between self and reality, and stereotyping.)

Summarizing our findings on the question of the effect of therapy method on the therapy relationship, we have not much to say. Apparently a leading method of therapy is more effective than a

9. Virtually all of the variables explored by *Core Group II* were analyzed with respect to the effect of changing the method of therapy on the dependent variables of client behavior. (See part I of the present report for a list of these variables.)

10. Baker constructed reliable and valid tests of these variables: (a) tendency to perceive indiscriminately, (b) tendency to resist analyzing one's problems, (c) unwillingness to accept reality about oneself, (d) tendency to overgeneralize, and (e) a test of stereotyping based on the ACE Inventory of Beliefs.

reflective one in reducing the tendency to make overgeneralizations. Apparently also, as would seem obvious, persons who leave therapy early show resistance to self-analysis. Apparently clients demonstrate more guarded behavior in leading therapy than they do in reflective, and possibly therapists are less critical of their clients' progress in a leading therapy. These facts would seem to be an elaboration of the obvious, but it is still desirable to check the obvious by research to find out whether it is really true. (The interaction of therapy method and therapist characteristics was a more productive variable, and will be discussed in the next section.)

EFFECT OF INTERACTIONS BETWEEN CLIENTS, THERAPISTS, OR METHOD, ON THE THERAPY RELATIONSHIP

Perhaps the most fruitful area of our research has been this final section, the effect of interactions between clients and therapists, or between therapists and method, on the therapy relationship. It was in *Core Group II* that it first became apparent to us that interaction was an important fact to consider, perhaps because the *Core Group II* studies were the first to make use of a double classification system of analysis of variance. This group of studies examined the question of whether the interaction of the therapist as an individual and the type of therapy he is employing affects the client-therapy relationship, and the results are somewhat affirmative. The way in which therapists used or molded a type of therapy had effects upon clients. Clients felt significantly more defensive or more positive in one type of therapy with individual therapists than did other clients for the same therapist in the second type of therapy. For some therapists, the increased defensive or positive feelings were in the leading treatment, while for

other therapists, the reactions elicited were greater in the reflective treatment.¹¹

The two most recent studies to be undertaken at Penn State are almost exclusively directed toward analysis of the effect of the interrelation between client factors and therapist factors on the therapy relationship itself. While neither is yet completed, I think it is permissible to discuss their character, and some of their aims at this time.

For the past 2½ years, I have been conducting an analysis of relationship factors in psychotherapy (11). The specific aims of this study are:

1. To conduct a pilot study and to construct a theoretical formulation about the nature of ten specific client-therapist relationship measures;
2. To carry out correlations of the interview-by-interview changes in counselor and client attitudes toward each other;
3. To analyze the ability of the counselor to perceive accurately the significant interview-by-interview changes in the client's attitudes toward the counselor; and
4. To complete a preliminary validation of seventeen measures of counselor and client interview-by-interview interaction in interpersonal attitudes.

Twenty cases of psychotherapy (conducted by this investigator) have been recorded in full on tapes. Only three cases continued for less than twenty interviews. After each interview, every

11. The clients in this study did *not* differ significantly as a result of the interaction effect in the extent of guarded, defensive, dependent, open, covertly resistive, or overtly resistive verbal behavior which they manifested. Nor did they differ as a result of the interaction effect in the extent of change they exhibited in maladjustment, dependence, defensiveness, positive attitudes toward the self, positive attitudes toward others, anxiety, or therapists evaluation of change as a result of therapy.

client completed a 200-item questionnaire relating to his interview-by-interview reactions to the therapy and/or therapist. This questionnaire contains 10 subscales of items believed to measure needs or motives such as "affiliation" or "fear of exposure." Half of the items are "positive" and half "negative."

After each interview, the therapist also filled out the above questionnaire, *as he believed the client probably* had filled it out, *i.e.*, the therapist's perception of the client's feelings toward therapy and therapist. In addition, after each interview the therapist filled out a 196-item questionnaire, designed to measure: (a) counselor feelings toward client and/or therapy, (b) counselor's estimate of client's feelings toward therapist, (c) counselor's estimate of client's progress in therapy, and (d) counselor's estimate of client's post-interview need-structure.

Further, after every fifth interview, the client responded to the Edward's Personal Preference Schedule. For the purpose of validating certain items on the new scales mentioned above, the client also early in therapy took the MMPI on one occasion.

The principal statistical devices being employed are tetrachoric correlation, Spearman rank-order correlation, phi coefficient, and factor analysis. The correlation of client and counselor interaction is based upon the intercorrelating of tests on the 200-item questionnaire as filled out by clients, and comparable tests on the 196-item questionnaire filled out by the therapist. The analysis of counselor perception of attitudes of the client is based on the correlation of the 200-item questionnaires filled out by client and by counselor-as-he-perceives-client-as-responding, respectively. The validation of the 17 interaction measures will be facilitated by a factor analysis of the twenty clients, together with a testing of the seventeen subscores, for correspond-

ence with the group factors obtained in the factor analysis.

It is planned to interrelate changes in the client-therapist relationship with such specific therapy techniques as certain types of interpretation, etc.

This is an exploratory study, and no attempt will be made to use it for proof except in the somewhat limited setting. However, it involves tremendously significant material in the derivation of inferential theory. It is not a severe limitation that only one therapist is involved because of the large sample of behaviors. However, the therapy is being recorded so that it is public information and can be judged by groups of judges with regard to certain variations. Thus the clients can be ranked for a variable amount of such factors as amount of domination by the therapist, amount of affection by the therapist, amount of directiveness, amount of countertransference, etc. No control group is needed and it would be pointless to use one. The study is designed essentially to produce hypotheses which can be tested in program research using large groups of counselors and clients.

Some of the measures which are feasible in this study are:

1. Correlation between first and second, first and third, etc., interviews to examine trends in attitudes.
2. Correlations between each interview for the client, and the same interview for the therapist.
3. After valid subscales have been factored, correlations of the degree of correspondence between client's self-estimates, and therapist's perceptions of client's estimates, from one interview to another.
4. Correspondence between therapist's estimates from interview to interview.
5. Correlation between client's estimates and therapist's estimates from one interview to another.

6. Degree of differentiation among clients.

7. Correspondence between the therapist's evaluations from one client to another.

8. Change in client scores, or "progress," in terms of a previously selected criterion of progress in the relationship factors, e.g., increasing independence, etc.

Some examples of hypotheses to be examined are:

(1) As counselor attitudes become more favorable, from interview to interview, client's attitudes also become more favorable, and vice versa

(2) Positive client-counselor relationships occur at points in counseling where the counseling interviews are judged to be less threatening to the client and/or counselor.

(3) Interpretations increase (a) client dependency and (b) client pos./neg feelings.

(4) Unreciprocated client negative feelings lead to more positive relationship.

I have available just a small fragment of results from this study which can be reported thus far. I have carried out a factor analysis of my twenty clients on their first Edwards test, and have computed some correlations between these factors and the rankings I have made on the clients on six different characteristics. There were four factors extracted on their Edwards, which I have labeled on the basis of the high-ranking Edwards sub-scores of each group, as follows:

- I. affiliation-nurturance
- II. succorance-exhibitionism
- III. deference-introception
- IV. achievement-autonomy

The six traits on which I ranked the twenty clients were: (a) strongest rapport, (b) liked best by counselor, (c) most dependent on counselor, (d) most successful therapy outcome, (e) most hostile to therapist or others, and (f)

most guarded in therapy. As you would assume, there are high intercorrelations between these different sets of rankings, only two of them being below $\pm .73$. Correlations of these rankings and the factors were high, but they are probably meaningless, since I was unable to demonstrate significant differences between the correlations for the groups of ten highest and ten lowest clients in each of the categories.

The final group of studies to be reported in this paper is one that I sometimes refer to as *Core Group IV*.

Together with three of my colleagues, Drs. Ford, Urban, and Ray, I am now engaged in a program of investigations (12) of a three-dimensional theory of interpersonal interactions in psychotherapy. We are presently working on the first of a series of studies designed to measure procedures developed to determine whether therapeutic interaction can be represented adequately by a conceptual model based on the dimensions of interpersonal control, affect, and disclosure.

This research group takes as its starting point several propositions concerning therapy research, and has endeavored to embody these within its research design.

1. That an adequate assessment of the therapy relationship must involve the investigation of *all three classes* of variables which hitherto have more usually been studied separately, *i.e.*, therapist variables, client variables, and variables descriptive of the particular relationship which evolves. We are therefore assessing these three variable-classes concomitantly.

2. That therapy behavior is most fruitfully considered as a sequence of interpersonal reaction patterns. It is thus proposed that an essential feature of psychotherapy is that it is a series of interpersonal transactions between two individuals who enter upon their relationship with a history of preferred habits

of personal interaction; that these habits of interaction are related to the type of interpersonal relationship which will subsequently evolve; and that the interpersonal relationship which does evolve will significantly influence the therapeutic outcome. It has followed that our formulations and hypotheses are couched in interpersonal terms, and that all of our measures constitute an attempt to assess characteristic interpersonal reaction patterns.

3. That the most effective predictor variables for the interpersonal transactions which constitute therapy are the characteristic or preferred habits of personal interaction which each member, therapist and client, brings with him to the threshold of therapy. As a result our research has been designed to secure measures of interpersonal reaction patterns of each therapy participant *antecedent to therapy*, and to trace the operation of these same reaction patterns within the therapy itself.

4. That significant dimensions must be employed for the adequate characterization of these interpersonal reaction patterns, dimensions which can serve several purposes, *viz.*, the adequate depiction of interpersonal reaction patterns both outside therapy with a multiplicity of individuals, as well as within the therapy relationship itself; and more particularly dimensions along which a high proportion of therapy verbalizations can be rated. We believe that we have found three such dimensions,—to us they appear quite promising, and to date we are rather encouraged in our discovery that there are few interpersonal behaviors occurring within therapy which cannot be represented in terms of them. We have labeled these *control*, *affect* and *disclosure*.

These dimensions can be briefly described:

(a) *Control*—reflecting the frequency with which interpersonal control of other

individuals is attempted; (this aspect of behavior has often been referred to under the label of dominance, or as directiveness when therapist behavior is described);

(b) *Affect*—reflecting the frequency with which negative/positive feelings are directed toward another person;

(c) *Disclosure*—reflecting the frequency with which an individual reports to another aspects of his experience and behavior hitherto inaccessible to the observation of others. In our work to date, these dimensions have been built into each of our measurement devices (cf. below on instruments).

5. That the *concomittant* assessment of interpersonal behavior on several dimensions at any one time, resulting in patterns or configurations, is an important methodological step in a study of therapy. We are proposing that the properties of a time-segment of any interpersonal transaction can be described in terms of its variation on each of these dimensions *simultaneously*. It is important in describing our procedure to emphasize that each verbalization that occurs within therapy is *simultaneously* rated on all three dimensions.

6. That the effective study of therapeutic events is dependent upon the adequate assessment of the content of verbalizations which occur. Procedures for content analysis have always been a thorny problem, and many investigations of therapy events have suffered for lack of an appropriate method of content assessment. This research group has evolved a procedure which appears to have some unique properties, and promises to be workable. We have come to refer to this procedure as *Time-line judgments*.

While developed independently, this conceptualization resembles the scheme developed at the Kaiser Foundation, and described in several articles and the recent book by Leary. In many ways the

system of Leary, *et al.*, is applicable to the aspects of interpersonal relationships with which we are concerned, but there are distinct differences in our system which we believe to be more appropriate to *individual* psychotherapy relationships.

Eight different devices have been constructed and administered to fourteen counselors and twenty-eight clients in the Division of Counseling and the Psychology Clinic. The cases are in progress, and the tests are being administered on from one to five separate occasions. The reliabilities and validities of the eight measures will be determined by use of correlations, and a number of subsequent factor analyses. It is also planned to administer the tests to three special populations, including psychotics, for validating purposes.

The measures being employed in this study consist of the following:

1. Structured Q-sort of attitudes toward the self. (Clients and counselors.)
2. Structured Q-sort of attitudes toward generalized-other (Clients and counselors.)
3. Structured Q-sort of attitudes toward the counselor (in case of the client.)
4. Structured Q-sort of attitudes toward the client (in case of the counselor.)
5. Rankings of therapists by supervisors on "Descriptive paragraphs"
6. Rankings of clients by themselves on the "Descriptive paragraphs"
7. Rankings of therapists by themselves on the "Descriptive paragraphs"
8. Analysis of selected interviews on "Interview Analysis Scheme."
9. Edwards Personal Preference Schedule (on the clients.)

Following the validation of the measures, new populations will be drawn from the Psychology Clinic and the Division of Counseling and a number of hypotheses will be examined. The following are typical of the hypotheses which will be drawn from the theory and then tested.

(1) Clients falling at the extremes of the 3 dimensions will show a smaller degree of change over the therapeutic interval.

(2) The more flexible, *i.e.*, the less committed a therapist is to any extreme position on the three dimensions, the greater the behavioral change in his client over a period of time.

(3) Resistance responses are strongest in clients in the dominant-hostile-overt classification.

(4) Interpretation is most used by counselors at the dominating end of the control continuum

(5) Therapists whose behavior is predominantly dominant and positive will show increasingly positive affect toward submissive clients as therapy progresses.

In conducting this research we have found ourselves confronted with several difficulties, which we are trying to find ways of solving. The first is the problem of analyzing the personal characteristics of experienced counselors. They are sufficiently sophisticated in psychological measurement to make it difficult to find valid ways of studying them. Perhaps the structured type of Q sort has been the most satisfactory. The other method with some validity is a measure of their actual therapy behavior, as rated by judges.

This leads to the second problem; finding or training enough adequately experienced judges. As soon as a graduate student is about at the point where he becomes a good judge he leaves the campus, and communication becomes difficult.

The third and largest problem is finding the money to underwrite the cost of large-scale investigations such as we are trying to pursue. So far we have not been able to locate a vacant seat on the abundant breast of the mother of psychological research, the Public Health Service.

It is somewhat more difficult to *summarize* findings regarding this question of interaction, because so much of it is still being studied. We have shown that the interaction of the therapist as an individual and the type of therapy he is employing affects the amount of positive and the amount of defensive attitudes which the clients display. This occurs with both leading and reflective methods of therapy, and it is the combination of therapist and method which produces the effect, a much more discernable effect, than that produced by either the therapist variables alone, or the variation in method alone.

We are at present examining the question of whether day-to-day changes in therapy relationship can be associated with certain therapeutic procedures such as interpretation, reflection, etc., or can be associated with variables of the personality of the client (MMPI, Edwards) or of the therapist (MMPI, Edwards) or of various aspects of the attitudes of the therapist toward the client, or the client toward the therapist. We suspect that most probably it is in the interaction of a number of these factors that we will find the key to some of the presently unknown characteristics of the therapy relationship.

Finally we are exploring whether it is possible to study the therapy relationship within the conceptual scheme of a three-dimensional system of the factors of control, affect, and unguardedness. If we find the latter to be true, we shall explore many hypotheses relative to the variance and interaction between these three dimensions of therapy inter-relationship.

REFERENCES

1. Ashby, J. D., Ford, D. H., Guernsey, B. G. Jr., & Guernsey, Louise F. The effects on clients of therapists administering a reflective and a leading type of psychotherapy. *Psychol. Monogr.*, 1957, 71, No. 453.
2. Aronson, M. A study of the relationships between certain counselor and client characteristics in client-centered therapy. In Snyder, W. U. (Ed.) *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Univer., 1953. Pp. 39-54.
3. Baker, E. The differential effects of two psychotherapeutic approaches on client perceptions. Unpublished doctoral dissertation, Pennsylvania State Univer., 1957.
4. Gallagher, J. J. The problem of escaping clients in non-directive counseling. In Snyder, W. U. (Ed.) *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Univer., 1953. Pp. 21-38.
5. Gillespie, J. F. Jr., Verbal signs of resistance in client-centered therapy. In Snyder, W. U. (Ed.) *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Univer., 1953. Pp. 105-119.
6. Kahn, R. K., Therapist discomfort in two psychotherapies. Unpublished doctoral dissertation, Pennsylvania State Univer., 1957.
7. Karmiol, E. The effect of the therapist's acceptance of therapeutic role on client-therapist relationship in a reflective and leading type of psychotherapy. Unpublished doctoral dissertation, Pennsylvania State Univer., 1957.
8. Mazurkiewicz, J. F. A comparison of the effect of a reflective and of a leading type of psychotherapy on client concept of self, of ideal, and of therapist. Unpublished doctoral dissertation, Pennsylvania State Univer., 1957.
9. Peterson, A. O. D., Snyder, W. U., Guthrie, G. M. & Ray, W. S., Therapist factors: an exploratory investigation of therapeutic biases. *J. counsel. Psychol.*, 1958, 5, 169-173.
10. Snyder, W. U. (Ed.) *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Univer., 1953.
11. Snyder, W. U. An analysis of relationship factors in psychotherapy. Study in progress.
12. Snyder, W. U., Ford, D. H., Urban, H. B., & Ray, W. S., Program investigations of a three-dimensional theory of interpersonal interactions in psychotherapy. Study in progress.

Discussion of Papers by Bordin and Snyder

MAURICE LORR, PH.D.

The reports of the research programs underway at Pennsylvania State University and the University of Michigan are impressive and stimulating. Certainly a program of related studies that utilizes earlier findings in the design of those that follow is more likely to yield pay dirt than single isolated efforts. Further, the analysis of therapeutic interactions, as we all know, is sufficiently tedious and time-consuming to demand sustained group effort.

There is much that is common to the two programs. Both agree that the key to the effects of psychotherapy on the patient is his relationship to the therapist. Both have examined in considerable detail not only the interactional aspects of the interpersonal relationship but have sought to identify characteristics of the client and the therapist that are relevant to therapy. Their general methodological approaches are quite similar. It is evident that Bordin and Snyder and their associates regard the problem of construct validation as important. In attempting to study the therapeutic process in an objective manner every investigator encounters the problem of objectifying and making more precise the basic terminology in the field. Concepts such as "transference," "warmth," and "defensiveness" are difficult to define operationally and their dimensionality is uncertain. Bordin seems to prefer scaling techniques to generate unidimensional variables. Snyder and his associates tend to turn to factor analysis for solutions. Both attempt to establish the validity of their constructs by reference to a wider network of personality and therapeutic theory.

There is another element of similarity also worthy of mention. The reports concern the research in which each is engaged. This would tend to reduce the number of references to results and activities of other researchers on the therapy relationship. It is singular, however, that neither paper has a single cross reference to the other. One would expect some interchange, some referral to the concepts developed by the other, to confirm or disconfirm. There is little in the body of the reports to indicate that this has occurred. The impression is given that these research centers have carried on their parallel efforts in comparative innocence of what the other was doing. We believe this type of isolation is characteristic of much of the work now being done in this area. Many psychologists and psychiatrists are seeking to isolate and measure identical variables. Yet the categories developed and procedures reported are sufficiently distinct to prevent direct interpretation and translation of one scientist's variables into those of the other. Perhaps one pay-off of this conference will be to bring researchers into closer agreement and into closer contact.

Resistance is a concept of importance in psychotherapy. Some regard the entire process of therapy as one of working through resistances as they occur. Is this a unitary concept? What relations does it have to the directiveness of the therapist or to depth of interpretation? If Gillespie's (2) paper is examined we find a definition in terms of resistance toward the therapist, toward the therapeutic process, and within the client. Bordin finds that resistiveness consists of two parts. First the patient's compliance

with or avoidance of his task to observe himself and report to the therapist. The second represents the patient's reactions to the therapist's interaction. A recent report by Ashby, Ford, et al. (1) appears to broaden Gillespie's concept of resistance and utilize concepts of client openness, guardedness, covert and overt resistance. Clearly there is considerable similarity here. Openness and guardedness seem to correspond to Bordin's factor of compliance to the task at hand. The overt and covert resistance indicators possibly correspond to Bordin's second factor. Obviously there is need to determine the extent to which the devices used to measure resistance tap common elements.

Bordin reports that depth of interpretation is related to resistance. Is depth of interpretation, which Fisher has shown to be virtually the same as "plausibility," similar to or identical with Gillespie's "directiveness"? For Gillespie, directiveness consists of inaccurate clarification of feeling, clarification of un verbalized feeling, and interpretation. It is probable that these are identical with the construct of depth of interpretation or plausibility.

This brief comparison of these two constructs illustrates the clouded and nebulous state of our knowledge of crucial variables in psychotherapy. Not only is there need for empirical studies of it but there is also an obvious need to check the conclusions of previous studies and to develop common measures of these constructs.

Bordin has reported in some detail the efforts that went into the examination of the concept of warmth. The perceived warmth of the therapist is found to consist of the therapist's degree of commitment, his effort to understand, and his spontaneity. Snyder does not refer specifically to the therapist's warmth. However, the measures used by Ashby, Ford, et al. (1) do seem to be measuring

closely similar variables. The therapist-characteristic variables chosen for study were ability to enter the phenomenological field of another, sympathetic interest, acceptance of others, and a number of other measures not of special interest here. The first was defined as interest in learning about the internal frames of reference of others and being able to see how others perceive and feel in terms of these internal frames of reference. The Intracception Scale, and the Nurturance Scale of the Edwards Personal Preference Schedule as well as a role playing test and a test of the ability to accept what the patient has to say with no need to judge were used as measures. It seems likely that the Ashby variables and the Bordin variables are approaching the same underlying group of constructs.

Snyder, as judged by the report, is more interested in the effects of a group of therapeutic techniques than Bordin although the latter has investigated single variables and their relationships. In the Ashby, Ford report contrast is made between reflective and leading therapies. Both are described as having in common the therapist's effort to create a warm, acceptant, understanding, noncritical climate. Each, however, has certain unique therapeutic features. One is based on the Rogerian approach and the other on the approaches of Dollard and Miller and Fromm-Reichmann. The independent variable is here the type of therapy. In Leading therapy it consists of such therapist responses as directive leads, interpretations, approval, encouragement, advice, information giving, and the like. In Reflective therapy it consists of restatement of content, reflection of feeling, nondirective leads, and nondirective structuring of responses.

Now this represents a degree of standardization and specification of the independent variable very considerably removed from the failure to specify what

should happen in the therapy hour. On the other hand, it is very many steps removed from the complete specification of therapist behavior as one might conceive in a Skinnerian experiment. Yet it is my feeling that we have to go much further in structuring, specifying and limiting the degree, intensity, and duration of the therapist's behaviors. It should be possible, perhaps through the application of learning theory, to define more narrowly and to manipulate experimentally, a series of independent therapist's variables

Where might research move in the next few years with respect to the study of the patient-therapist relationship and the process of therapy? What steps might be taken to further research?

The following are some possible approaches:

1. There is presently available a half dozen or more systems for the rating, coding, recording, and analysis of what goes on in the therapeutic interview. An effort should be made to relate these alternative systems. Comparisons based on common interview data might make more evident common elements, relative feasibility and reliability, as well as the unique features of each system.

2. As many have indicated, there are a very large number of patient variables, therapist variables, and interaction variables that appear to have some relevance for therapy. Many of the fuzzy duplicates can be eliminated by operational definition and scaling. Acceptance of a common set of scales, a common language as it were, for use in research would in itself be helpful and something of an achievement. However, this procedure would not assure their unidimensionality. To reduce the number to a minimum set and to provide a workable framework some such technique as factor analysis is needed. At present the procedure in objectification and quantifica-

tion of variables is from simple categorization, to scaled variable, to Q-sort or questionnaire. Typically the goal is to develop reliable measures of various constructs. There is need to utilize a representative group of the better developed variables on patient populations. A series of factorial studies would be helpful in constructing firmer formulations of what happens to patients and therapists in the therapeutic relationship.

3. In a process study the investigator takes numerous observations sequentially. In an outcome study observations are typically made only before and after therapy. Let us assume that the process of therapy is a lawful stable series of events. It should then be possible to isolate and identify time sequences of behavior. Such behavior pattern sequences or sequence analysis might prove to be unusually illuminating for therapy. Some work along these lines has already been done in other fields under the label of spectral analysis.

4. Currently the therapeutic process is being intensively studied through the verbal interaction of patient and therapist. There is need to secure simultaneous records of verbalizations, psychophysiological correlates, and of patient-therapist body movements. The framework of linguistic analysis being developed by the anthropologists to study language and paralinguages should be explored. My own hesitation is that such a minute, tedious, and microscopic structural analysis will lead into blind alleys. I find it difficult to see how such an analysis is justified when therapeutic changes as ordinarily observed are so slow. Yet, the method seems worthy of trial.

5. It would be a considerable benefit to all researchers if a library of tapes could be established. Contained in this library, perhaps supported by USPH funds, should be a representative group of therapy tapes on a variety of patients

and clients treated by psychoanalysts, nondirectivists, etc. These tape sequences could be used for checking and cross-checking coding and measurement schemes. They could also form the basis for studies of the comparative effects of various specific therapeutic techniques. The net result would be to make available for public inspection a sizable body of data on a variety of therapies by their most skilled exponents.

REFERENCES

1. Ashby, J. D., Ford, D. H., Guerney, B. G., Jr., & Guerney, L. F. Effects on clients of a reflective and a leading type of psychotherapy. *Psychol. Monogr.*, 1957, 71, No. 24 (Whole No. 453), 1-32.
2. Gillespie, J. F., Jr. Verbal signs of resistance in client-centered therapy. In Snyder, W. U. (Ed.) *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Univer., 1953. Pp. 105-119.

Therapist-Patient Relationship

DR. SASLOW: I would like to ask a question which I meant to ask ever since reading everything of Dr. Bordin's. I have been bothered about the way in which you rank-order depth of interpretations in relation to the discrepancy between what is said by the interpreter and the lack of understanding or awareness of the interpretation on the part of the subject. I can imagine all sorts of reasons for major discrepancies there, but why should that be the same thing as rank-ordering *depth* of interpretation? Interpretation might be a thousand miles wide of the mark for the subject, and hence he may be completely incapable of responding to it; just such an interpretation is rated by you as of maximal "depth." You must have some way of dealing with this, but it is not clear to me just how.

DR. STRUPP: There is a more general point, which was suggested in Dr. Bordin's paper. He says that many of the most significant aspects of the process of psychotherapy have been left unexplored and uninvestigated, or investigated only superficially. For example, transference is one of those key variables. Anyone who works in the analytic tradition and observes patients in therapy knows that patients distort the therapeutic situation and the therapist in particular in more or less specific ways. I would like to raise the question why it is that we find it so difficult to come to grips experimentally or in a naturalistic way with these key variables. One of the reasons obviously is that it is so extremely diffi-

cult to pin them down and to devise operational measures. I am wondering whether we are not often drawn in the direction of looking for relatively easily quantifiable things rather than those that are crucial and significant. This applies not only to misperceptions on the patient's part, which we consider to be very important—one of the major goals of therapy is to dispel these distortions and misperceptions, at least in analytically oriented therapy,—but also to distortions and misperceptions on the therapist's part. The latter have been grouped under the heading of countertransference.

All of these things we are very poorly equipped at the present time to deal with, but I also feel that we could make more efforts along those lines in trying to come to grips with what is doubtlessly very important in the doctor-patient relationship. I wanted to throw this out for discussion.

DR. BORDIN: I have some feeling that I ought to first start by commenting on Dr. Lorr's comments.

I think my first reaction is to say that I feel that I agree very thoroughly with most of the comments that have been made. I think you will find in my paper that I, too, am calling for this convergence and a movement toward employing the same kinds of measurements.

We ought to be using similar measures rather than measures that are almost similar but different enough so that you hesitate to try to throw the studies together.

I guess I at least have to question what I think is a methodological issue. It

* Abridged. See Editors' note, page 49.

seems to me that Dr. Lorr expresses a preference for multiple factor analysis as a way of identifying unidimensional attributes—continua—as compared to what would be a nonparametric method, such as the techniques that Hayes has developed, which are in many ways equivalent to factor analysis. I don't know that I would be prepared to argue the methodological grounds. My foundation is much too weak here. I would have to leave the field to the methodologists. But it seems to me that this seems to be the issue. It is my impression as a close observer of the scene that it is at least moot.

The multiple factor approach which Dr. Lorr calls for certainly no one can quarrel with. Our problem is to take into account as many sources of variance in behavior as we are aware of as significant. I just think that we should not underestimate the difficulties in setting up such a design.

I think we have to live with a certain amount of experimental error, a variation that refers to important independent factors in the therapeutic process, but which for the moment we allow to be a source of error in our experimental design. Sometimes we just have to live with it. I certainly feel that we should, wherever feasible, move toward multiple factor designs. I don't think there is any argument on that.

Then finally Dr. Saslow's question, and I think I can answer it in one minute: this is a defect in our measurement. We have just thrown that into the pot, acting as though the therapist communications are always accurate.

DR. DITTMANN: I take a somewhat proprietary view toward this scale since I was on the project at the time it was being developed. The scale leaves out the possibility that the therapist might be incorrect in his interpretation. It tackles just one meaning of depth among all possible meanings, the degree to which

the material being interpreted is removed from the patient's current awareness, as opposed to the degree to which it refers to early genetic events or whatever. I am not sure that leaving out "correctness" is a defect in the scale, but here is the way that the scale was applied: judges were allowed to circle any judgment that they made if they felt the scale did not apply to the statement being rated. One of the instances where the scale might not apply is when the therapist, in the judge's opinion, misses the boat completely. As a judge, you may circle any individual therapist's response as being unratable—because the therapist and patient are exchanging information at this point, or because the therapist is so far off the boat that the scale of depth of interpretation is not meaningful. Relevance is not an issue on the scale itself.

DR. S!R! P?': May I make one point on the same remark? We have worked with a similar scale that was not restricted entirely to interpretations as Bordin's scale was. I would support the point of view that was presented. It does not make too much difference in our experience whether depth or inference is defined in relation to what purports to be the patient's understanding or whether it is defined in more absolute terms. There seems to be a fair degree of rater agreement on the relative depth of the therapist's communication. These judgments can be made.

DR. FRANK: First I would like to ask one question just because I am sure I misunderstood it. Dr. Bordin's work is so condensed and carefully thought through that if you are not in the framework, you don't always get it. I thought I understood him to say that one criterion of a good measure is that people trained in the theory would use the measure better than naive people. This may be all wrong. Isn't a good measure one that anyone can use? A good rule

can be used by people who are trained and not trained. I think I missed the point here.

The other comment is more positive. I am very fond of this concept of commitment. I think it is probably a very important notion. If one takes the view that one of the attitudes of the therapist that produces changes in patients is the attitude that conveys to the patient that this man is really trying to help him, this relieves some of the dread or anxiety that Dr. Whitehorn spoke about in his dinner address. The therapist's attitude of commitment brings relief and sets in motion changes in attitude so that the patient feels differently about his problem.

In connection with this it seems to be quite close to the kind of data that Dr. Whitehorn and Dr. Betz are turning up with respect to doctors who do well and doctors who do poorly with schizophrenic patients. They took 100 consecutive schizophrenics at Phipps over a period of a decade, and found that certain doctors consistently got much better results than others. The doctors who did well had the attitude of what they call "active personal participation," which seems to be very close to this notion of commitment and spontaneity. This is different from what you seem to imply when you talk of parents or teachers offering help. It is not merely offering the help, but offering it in such a way that it is perceived as help. For example, a therapist might offer help by telling the patient what to do, but it may not get across to the patient that this man was really trying to help.

DR. BORDIN: That is a very important problem and we are just feeling our way through it. Primarily, I think, as you are suggesting, it is more the expectations that the person has in the relationship. What does he see or expect of this other person? You cannot separate

the expectations from at least some input of behavior on the part of the therapist. I am talking now about the expectations that are realistically determined by the therapist behavior as compared to built-in expectations. I am trying to separate those out. To some extent I am suggesting that these need to be separated out. Different patients will bring different kinds of expectations and anxieties.

You were emphasizing one side of it as though commitment was always good and would always contribute positively to therapy. I think we are seeing that under certain conditions commitment might contribute negatively to therapy. As a matter of fact, this is one of the things we are trying to avoid, that is, the formulation of attributes of the therapeutic process that have built-in answers on a kind of a value basis. More of this is good and less of it is bad. We tend to be suspicious if we find ourselves thinking in those terms. I think the reason is that we assume that there is enough variability in personal needs, and consequently variability in the conditions that will lead to change, so that the high end of any attribute is not always going to contribute to change.

DR. FRANK: Let me see if I am thinking in the same way as Dr. Bordin. In my terms, what he is saying is that the main thing that counts is the patient's expectancy of help, and this will be aroused in different patients by different kinds of behavior on the part of the therapist. So commitment will arouse expectation of help in some patients, but not in others.

DR. BORDIN: There will certainly be that factor in there. Also, the offer of help becomes something that the patient reacts to. We are trying to predict how the patient will react to offers of help.

DR. FRANK: I am raising the question, don't you have to specify a little more the way the help is offered? I think

you do in these scales. An important distinction may be whether or not you offer the help in a committed way.

DR. BORDIN: Yes. There certainly is a problem there that might involve the issue of spontaneity and this we have been concerned about.

DR. FRANK: Might I also ask about the scale question so you don't forget it?

DR. BORDIN: I wanted to forget that one. This is the question of why do we demand that the person with more training be able to do it? I am not sure that we make this as a general assumption. We think that there will be kinds of observations where you expect that experience or some other qualification of the observer will enter in. We thought in the case of depth of interpretation this was likely to be the case. It seems to be a highly demanding kind of observation to be able to pick up and to understand the patient's current level of awareness.

DR. SEEMAN: I do want to go back to the question of depth. This is certainly a commonplace observation and yet terribly important, I think, and that is that one needs to specify his terms; depth is a very good example, Depth A, Depth B, Depth C, and so on, until you have exhausted the range of definitions of depth, so that we really understand what it is we are talking about. We can define depth in different theories of therapy and even within any given theory of therapy we can define depth according to several quite legitimate, but different criteria. So we will get hopelessly confused about the terms unless we do specify the criterion. I think that is one of the values of your study, that you specify the criterion. Maybe you should have specified the criteria of the term "interpretation," too, because you can have Interpretation A, B, C, and D. I don't think it is getting too picky to require that you do this. It helps to solve

some of the problems in communication, because I think one could derive various kinds of interpretations that would be differentiated when you really got down and looked at it.

DR. DITTMANN: In the study of the dimensionality of this variable depth, one of the things that turned out was that there is indeed a confusion on the part of many judges. The first factor that turned up was what we hoped it would be, namely, our definition of depth. Another factor that probably turned up, as near as we could judge, was something that is probably confused with plausibility, or how much the patient will understand what the therapist has said, or how upset the patient might be by this, or something like that.

DR. SHAKOW: I have two questions. These are really general questions for all of us but I think Bill Snyder's presentation brings them to the fore. One has to do with the expertness of therapists and what problems are raised by using students in training for therapeutic research. I am sure such a practice brings in all kinds of problems. We have touched on them at one time or another. I think that we run into this problem very seriously in studies carried out in psychological settings. It is even a problem in psychiatric settings, but there one is more frequently likely to get expert persons involved in therapy because the institutions are set up in such a way that therapeutic activity is a primary function. The exceptions occur when young residents are used. You have the example of the Menninger Clinic study: Here are a group of expert people who have to carry out the therapy as part of their job. When you use students, however, aren't there many problems that one runs into? I think we have to face this issue, because I think many of the research problems that we run into probably grow out of this particular practice problem.

I am much concerned, too, about what happens when you take inexpert people and instruct them in using certain special techniques of therapy. We know how difficult it is under any circumstances for a person who is in the process of learning a new kind of complicated operation. Aren't you creating additional problems if you at this time also tell him to do something quite different. You brought out very clearly what a problem you had. They went ahead and did what they pleased anyway.

You raised many questions about the kinds of confusion that have gotten in, the degrees of contamination that may have entered into the picture. After all, you have given your therapist one set of instructions, but he goes ahead and instructs himself in quite another way. Under these circumstances how clean is the operation that is actually carried out?

The other question that I have has to do with the problem of the inter-therapy periods. This again is quite interesting when compared with what is done in the Miller study. There they don't do anything to the patient in between. They only do something at the beginning and at the end of the therapy. We want to know what is happening during the process of therapy. So we intersperse questionnaires, we intersperse Q-sorts, or carry out other kinds of operations. When we do we rarely ask ourselves what is the effect of this kind of an operation on the therapeutic operation which we are carrying out.

You might be interested, in this context, in a study that Parloff from our laboratory did. He and Iflund carried out a study on the transmission of values from therapist to patient during psychotherapy. As part of their study they interspersed between sessions—these were daily sessions, as I remember—a Q-sort by the patient relating to the importance of the values discussed that particular

day. It turned out that one of the patients actually thought that this Q-sort was the therapy, and the other operation was merely a secondary aspect of the study. Perhaps it was more therapeutic actually for the patient than the therapy itself. I have to admit that these were two paranoid patients. And this one was the more disturbed of the two.

What might be the contaminating effects of such interspersed questioning or testing? This is a problem that has troubled me a great deal and I have raised the question with myself many times. One hates to let the possibility for collecting relevant data go by. One feels one is missing an awful lot. At the same time you wonder, if you bring up material related to the therapy again, are you not really modifying the process that you are studying?

DR. SASLOW: The first question I have (for Snyder) concerns the nature of what Bill Snyder calls the directive leads he gives to his counsellors. I would like to put this question in a particular context. It seems to me that the example that Carl Rogers gave of some observations of his own this morning indicates that to specify that a counsellor reflect or clarify a statement may describe only a very small part of what is perceived as important by the person on the receiving end of therapy. You remember, Dr. Snyder, what Carl pointed out was that while he would be hesitating with a particularly difficult statement that he was trying to reflect and clarify, etc.—ready to make a succinct summary of a complex communication in, so far as he knew, his best nonjudgmental manner—the patient was regarding the very hesitancy as a threat to him and wondering if a sword of Damocles was going to be let loose on his head

I submit you do not know enough of what is going to happen when all you specify is reflecting, being non-directive,

directing, clarifying, or any other terms you mention. It looks as if we have to complicate our design by specifying other features of the interviewer's behavior. This, for example, may have to be whether you hesitate or not, to return to the Rogers example. Perhaps what we shall finally find we have to do is to specify formal characteristics of the speech behavior of the interviewer that would go along with reflecting and clarifying, and in other experiments to specify similar formal characteristics but with words which are directive or leading words. We may have to experiment with all of these variables. They are related to each other in ways completely unknown if you leave important ones out.

The second question has to do with your remarks about eclecticism. I too happen to feel strongly, as you do, that this has become a dirty word, and quite unjustifiably so. What you describe is very likely to be a picture of the future with regard to how therapy will go on, despite the fact that more narrow interpretations and orientations have their usefulness for development and testing of theory. I think it will turn out in the future that one cannot be completely non-directive, reflective and clarifying. There are other variables of importance. It will also turn out that learning theory is of major importance for psychotherapy, even though Freud had no learning theory. We shall in general be forced outside of any one of the existing personality frameworks, as too narrow.

I submit that in the future most therapy will be eclectic in this sense. I think it is time to protest against and to stop this devaluing of a word which has a best sense, not only a worst.

DR. SNYDER: I would first seize the opportunity to just say a word or two about my comments about Dr. Lorr's material. I was delighted and pleased with what Dr. Lorr said, and he let us

off very easily, I thought. I agree with almost everything he said. I wanted to mention that there was communication between Michigan and Penn State.

This communication, despite the fact that we are the best of friends, is slow, because of publication lag. Ed is not going to sit down and write me a long letter about what he is doing. I wonder if things like the Counseling Center Discussion Paper at Chicago are not going to be very helpful for this sort of thing; maybe some of that informal communication ought to get going among groups of people who are highly interested in one subject, rather than having everybody sitting around tightly pressing to himself his secret ideas for fear of somebody stealing them.

I agree with Dr. Bordin about the fact that both nonparametric and parametric and factorial approaches are feasible and necessary. There are times when it is just not possible to get into a factorial design and you have to do something about it. Now I will get down to the questions and not try to avoid them any longer.

Dr. Shakow's question: The question of the expertness of the therapist and the dangers of using student therapists is certainly a problem that we are more and more becoming aware of. It is our present intention in the last study mentioned here to try to use much more expert therapists, people who have their Ph.D.'s, and have done several hundred hours of therapeutic counseling before they are used in this study. In some of the preliminary parts we are going to choose from among the inexpert therapists and find the less inexpert of them. Some of our people, although they are students, come to us with a good bit of experience. One was a lay analyst, for instance, from the Washington Institute. Although we have some terribly confused students, we also have some students who

are beginning to know what they are doing. I think you are quite right, it is our job to make as much use as we can of the more expert, because it is a very real problem.

You asked about the effect of instructing inexpert people in using techniques I don't know that I quite follow the question. I know what you are driving at. This is very difficult. It does contaminate and does complicate, and you have to have some kind of check on the degree of the success of your instruction. You can't take their say-so that they understand. You have to have some reliability checks of what they are doing. That is why we are so careful to record what they are doing, and to go over it, and have judges evaluate, and we get judge agreement as to whether it is really what they claim it is. It still does not solve the problem.

The second question was on the effect of interspersing questionnaires and other data-collecting procedures during therapy; this is certainly a problem for all of us. To what extent does your observation of the experimental animal change his behavior? It has always been true. I will propose this defense. You proposed a defense yourself yesterday when you said you know what it was like 15 or 20 years ago when anybody introduced a recording machine. Good Lord, they were ruining therapy! Now they take a machine for granted. Everybody does.

DR. SHAKOW: This is a somewhat different problem, though. I think you have to face that issue, that is, if you are going to get your primary data uncontaminated. The question is, once you have your therapy started, what happens? The whole problem of the attenuation of the transference in the relationship, or whatever else you want to call it, arises. If you ask questions about what is going on in therapy—or somebody else

comes along and asks questions—and the patient has brought into awareness problems about which he might not be aware or if you ask him all kinds of personality questions which he is not ready to answer yet—what happens to your succeeding therapy hours? I don't know. I am raising the question: What does such activity do to the therapy process?

DR. SNYDER. I will admit it does something. For example, a client will say, regarding an item on the questionnaire, "I would like to spend all of my time with the therapist. I didn't use to feel that way and I suddenly noticed I started to feel that way." This is affecting it a little. But, we are keeping it constant. We can say that about it. It is being controlled in that sense. Whatever is happening is happening to everybody and is happening every time. What its effect is we could measure by a control setup but we have not felt at present that we want to spend time on that particular problem.

DR. SHAKOW: You understand I was not raising it only about your study. It is a problem for all of us in the field and I think we have to face it and try to find out what some of the possible answers are to this question.

DR. MATARAZZO: Mr. Chairman, I can contain myself no longer. I think this is an important question that Dr. Shakow has raised.

The problem you have raised, Dave, is as important in every field of scientific endeavor that I know of. For example, Heisenberg very clearly stated the principle (of uncertainty, I believe) that by the very fact of attempting to measure in physics we distort the data (i.e., the effect of the interaction between the physicist-observer and the situation observed), and this is a fact that is no less true in psychotherapy. We have to live with it in physics, in economics research, in psychotherapy, etc. As we

attempt to observe what is going on, we distort and change this in certain ways. They have learned to surmount this fundamental obstacle in physics, economics and other fields. They do this by accepting the fact that the observer by his very act of observing will distort the thing being observed, and so manipulate that which they are looking at in ways that are under their control to see what the relative changes are. When a government agency in Washington takes a look at unemployment or some other economic indicator in this country, it changes the behavior (unemployment) picture by so taking a close look at it. We can live with that, and we can record the effects that take place when we are aware of this as an experimental variable and take it into account by use of some kind of control, or by varying it so much in this case, and so much more in the next case.

I think the point you raise is a good one. It would not scare me as a researcher.

DR. CARTWRIGHT: I would like to take off from this question of whether or not one is modifying the process that one is studying. What I understood certain other members of the group to be saying, such as, for instance, Dr. Robbins, was something quite different from this, namely, that every individual is working in his own habitat, creating his own explicit behavioral environment for the subject that he is interacting with, that he has his own personally derived set of reinforcers, and his own systematically developed set of techniques; each has different numbers of interviews, different biases and different kinds of people he is working with; and there really is not too much comparability.

I don't think anybody has actually said this, but I do believe there are some people here who would say that if the average number of interviews was only

30, what relevance do any of those results have for our conditions where what we have is an average number of interviews of 610.

This is the question: Are we taking different aspects of one process? We can concern ourselves with problems of mucking up our observations by actually observing and so on, and affecting the process we are studying. Or are there perhaps *many different kinds of processes*, in many different ways, not only *substantively* different kinds of process, but maybe even there are *methodologically* different kinds of processes, namely: about the therapist and what he is doing over time; or, processes about the patient—Rogers' process scale of therapy is all about the client or the patient. It has nothing to do with the therapist.

There are some sorts of processes, such as interactions between patient and therapist on various occasions, which really do not have anything to do with outcome. I believe Dr. Bordin was saying he is not interested in outcomes, and whether anybody ever gets well or anything like that. It is just a matter of what these two people are doing to each other. It is a completely different kind of process again.

DR. LACEY: Heisenberg's principle of indeterminacy has become the new kind of homeostasis. Heisenberg formulated a principle concerning the impossibility of simultaneously measuring velocity and position of particles. Now this has suddenly become generalized into a new principle that every time we observe, we are changing the very process we are observing. I don't think it has anything to do with Heisenberg. This is a kind of convenient way of linking ourselves up with physics.

I think the issue is different. I would like to ask Dave what he means when he says, "Are we interfering with the process of therapy?" Whose process of

therapy? Are you somehow thinking of some standard, accepted, effective, defined therapy, let us say the 612 hours on the psychoanalytical couch, as therapy? Suppose it had all started in a different way. Suppose Freud had said that every eighth analytic hour we should sit the patient down in a new position and ask him to tell us what he can now recall about his infantile life. I am deliberately picking something ridiculous. Could we now say that we were interfering with the process of therapy? I don't think so. I think the defensible answer is this: In the present state of our precise knowledge of this field, I think we have got the right to set up any standard mode of operation we wish. If this includes interpolated Q-sorts, fine. If the investigator is honest, he reports, as these investigators have, that what has happened with the Q-sort material itself becomes occasion for discussion in therapy. I don't think there is any question of interfering with the process we are observing because the process we are observing is not real yet in the sense that it does not have rationally defined and accepted boundaries.

DR. WALLERSTEIN: I would like to go back to the problem stated by Dr. Shakow and restate it as a quantitative problem. To take the example of physics, every time you put a thermometer in a beaker of water you alter the temperature of the water in measuring it. You alter it so little that it makes no difference in the gross temperature you want, which is the nearest half degree. In psychology we have the same problem, but multiplied many fold and at several levels. Several levels get involved. One thing is what happens when you make a recording. We made clear in our presentation yesterday that in our project we don't do any recording. It does not mean we are committed in theory against it. At some point we may

do it. We feel at the point at which we begin to do recording, we will have introduced a systematic variable into the setting. It is different therapy than what we did before. At the point at which we do recording, we will want to study what this variable is

I think Dr. Shakow is committed to the proposition, and many people are, that you gain so much by doing recordings that it is worth adding this extra complication and taking it into account. I think there is yet another quantitative level. You not only make a recording which a patient gets used to and which you hope is unobtrusive, but when you introduce questionnaires after each session, you are introducing a second therapeutic hour in between what you call the therapeutic hours. At least a second hour in which the patient is thinking about the same issues and reacting to them. When you say that is a constant, I don't think you have covered the field. The questions may be constant, but the patient's reaction to those questions each day is not a constant, because different things are stirred up. This again is a much more major variable that has been introduced and which has to be systematically studied for its effect on the process, to have any relevance to the process these other people are studying who don't have that. I think that has to be made explicitly the focus of part of your research effort, so that others of us can compare it with what we are doing.

DR. SHAKOW: Dr. Wallerstein has said essentially what I want to say. I agree with John Lacey, the important thing is that you be clear about what you are studying. I am interested in studying the psychotherapy process. So I study the successive hours of psychotherapy as they come along. If I introduce procedures in between which get at the same kind of material that the psychotherapy process presumably gets at, this may

arouse all kinds of complications. These are not just additional observations. Observations are quite different from introducing techniques which really could have serious contaminating effects on the procedure one is studying.

DR. ROGERS: I just want to make one comment which really started with Dr. Saslow's remark about the difficulty and complexity in specifying the factors in the relationship. I think that part of that difficulty is due to the fact that we are still in the grips of a stimulus-response psychology, and we try to specify the stimulus. We try to specify what the therapist behavior shall be in the relationship and that is the stimulus in the situation. But I think, as several of you have suggested in your discussion, that it is always the *perception* of what the therapist is doing that is the crucial factor.

Take an element like empathy; certainly it is important that the therapist endeavor to understand; yet the crucial thing in the experience is not whether he is intending to understand. It is whether the client or patient *perceives* him as intending to understand.

I believe from a research point of view we might advance research in the therapeutic relationship by continually focusing not on what the *therapist* thinks he is doing or how we measure *him*, but on how what he is doing is *perceived* by the other individual. For me that resolves a number of the problems that otherwise are quite different.

Take what Bordin calls the possibility of being over-understanding in a parent-child relationship. The point is that such behavior is not perceived as over-empathy by the child. It is perceived as "meddling in my business," which is quite a different matter.

I am pleased at the broadening element in Bordin's study of the relationship, on the basis that a helping relationship is a

helping relationship, whether we call it psychotherapy or a parent-child relationship or friend-friend relationship. It opens up many avenues for easier study in which I think we can investigate some of the basic variables of therapy without always having to investigate them in the therapeutic relationship, which after all is not the easiest situation to get into the laboratory.

DR. HUNT: I want to come back to Dave Shakow's concern with the inexpert therapist and his concern with the introduction of and use of various therapeutic approaches by individual therapists. I would like to look at this kind of research from the standpoint of ethics. Thus far, no one has mentioned the word *ethics* in this connection, but I thought I saw the word looming up behind Dave's head as he was talking.

If you really know what is good for clients or patients, and more specifically, if you really know that inexpert therapists are, indeed, bad for patients, then I would agree that such a practice of using inexperienced therapists is indefensible even for research purposes.

Similarly, if you really know that having therapists who are trained to do "leading therapy" attempt to do "reflective" therapy is bad for patients, I would agree that such a procedure is indefensible. Unfortunately, the topic of this session is "the process of psychotherapy." And, as I have posed this issue, the key word is *process*, for nobody has talked about the relationship between process and outcome, or between process and what happens to patients. If process, and all the other aspects of psychotherapy, are continuously considered in relation to what happens to patients, I believe the issue gets focused in another way.

For instance, I defy anybody here to indicate existing evidence that will convince this group, that the inexpert psychotherapist is any worse for the patient

than the expert psychotherapist. At least, I know of no published data showing that the inexperienced gets less improvement in his patients than the expert. Similarly, although I may feel that asking a psychotherapist to utilize procedures which do not come "natural" may damage his psychotherapeutic effects, I know of no evidence to this effect, so I really do not know. So long as the issue is unclear, then I believe ethical considerations demand that the issue be clarified. Thus, when such issues are in doubt, and the investigator and the practitioner do not know what the effects on outcome may be, the ethical obligation is to permit all the variations of practice there are and then to determine their effects.

Furthermore, I believe that ethical obligations demand that the practitioner with strong feelings about what is appropriate and best for the patient, but without objective evidence for his feelings, should put his feelings to empirical test. He is obligated to compare the effects of his procedures on patients with the effects of other procedures for which other practitioners have strong predilections. Moreover, so long as the predilections are not based on intersubjective evidence, tolerance is demanded of each practitioner for the felt convictions of others.

Still another point is relevant here, namely, the variations in convictions about practice and what is good for patients provide the investigator with independent variables in process and therapist behavior to be tested against therapeutic outcome.

DR. SNYDER: I wanted to comment on this point; some of these student therapists charged us with being unethical by demanding that they use one method, and we always told them, "How can you prove that you are right or wrong?" But more interesting than this, and there are no published results on this one, we saw

very little evidence that they did any better on the one they liked better than on the one they did not like as well. The most amusing thing was that some of the clients would come in and say, "I want non-directive therapy," or "I don't want non-directive therapy," but we gave them whatever random order called for; they would be either happy or unhappy with what they got, but it made no difference which they got, they had their attitudes about it anyway.

DR. ROBBINS: I think my remarks relate to what I partially stated this morning, and what Dave Shakow brought up further. Perhaps an anecdote will be more helpful than being theoretical at this point.

I certainly am an example of an individual who is largely committed to a point of view, with right or wrong, a feeling of confidence that a certain therapeutic approach is useful and more useful than others. Quite frankly, after I received all the papers for the conference and did my homework, I came here with many misgivings about what we could accomplish, but I come out with a different idea after sitting here for two days.

I do not think it is important for us at this moment to be discussing or concerned about ethics. I think the ethics is a greater ethics of the scientific point of view. I think the important point that Dave is trying to make and I tried to make this morning is simply this. I don't think there has been a single research in this series of presentations that does not have a right to exist, whether you are using an unqualified student or a scrub woman with a galvanometer on her back. I think what is important since we are all at the moment trying to understand and ultimately benefit people, ourselves included, is what are these processes that are so abstruse and hard to get at. We do have to recognize the noncomparability of our methods. We

have to recognize how these processes are affected by the process of observation that we are employing. We also have to ask ourselves the question which was just briefly touched on this morning: what is the relevance of the measuring devices which we are using to the process we are trying to study, and does the measuring device change the process so it becomes an irrelevant experiment?

Some of the experiments I don't know enough about to judge. I only know this. In the exchange that we have had in these two days, as I heard more about these researches every one of them has become more relevant from the point of view of the guy who is doing it. My own attitude has become less judgmental.

I don't find myself in conflict with myself in sitting here and growing to love all of you. Better yet, gradually I am even gaining some respect for you. I think this is a very important and worthwhile thing.

I would like to go back to the scientific issue. I think there are some wonderful experiments going on here. Each will add, though they are as varied as Dr. Cartwright said. The important issue is to know what we are doing when we are doing something different, and not try to translate too literally every phase of every research into some single set of dimensions. There are too many variables and too many processes for us to do that at this time.

Research Problems in Psychotherapy

MORRIS B. PARLOFF, PH.D. AND ELI A. RUBINSTEIN, PH.D.

Implicit in the organization of the conference was the assumption that by bringing together a varied group of researchers, each actively engaged in the investigation of some aspect of psychotherapy, problems basic to the field would be brought into clearer focus. It was anticipated that the conference would not formulate solutions to these problems, but might rather provide an opportunity for the participants to begin to communicate in a more meaningful way. What was achieved was a remarkably free discussion of some significant issues with which these participants were preoccupied.

It is the aim of this summary chapter to underscore the central issues developed in 1) the papers read, 2) the prepared remarks of the four discussants, and 3) the spontaneous comments of the participants. This chapter is not an attempt to review systematically the content of each paper, but is an effort at an orderly presentation of the major issues considered at the conference.

Since the papers and formal discussion are reproduced here in full while the report of informal discussion has been abridged, this chapter draws heavily on the opinions expressed in the discussion and summary periods

Special mention should be made of the final morning summary discussion period, which has not been included in this publication. During this last morning the entire period was devoted to a spontaneous and informal group discussion of the prepared papers and the previous discussion periods. This last session produced a most useful exchange of comments. Valuable as this discussion was to the conferees, the verbatim transcript reveals that something is lost in the "written translation." The non-verbal cues, the quick transition from one construction to the next—based on the intimate interaction of the previous two days—made the editing of this part of the conference discussion most difficult. It was decided, therefore, to omit the full text of these summary proceedings from this volume but to incorporate the salient points of that discussion in this chapter.¹

1. Special acknowledgment is made of the many points derived from the evaluative statements of Drs. David Hamburg, Rosalind Cartwright, Joseph Matarazzo and Hans Strupp in their individual summaries of each of the four topic areas on April 12, 1958.

In the course of the conference it became apparent that participants were basing their comments on a variety of theoretical assumptions, variables, designs, and techniques. To facilitate presentation, this varied material is considered here according to its implications for: Research Goals, Methods, and Selection of Variables.

RESEARCH GOALS

Clearly, different aims may require different techniques. Although it is equally true that the same aims may be

approached by quite different techniques, the emphasis of the conference lay on specification of different goals rather than on questioning the appropriateness of the particular methods used to achieve them.

For the most part the researchers agreed that their ultimate goal was to identify the relationships among specific significant personality variables of therapist and patient, specific interventions, specific circumstances, and specific results. Few were rash enough to tackle this formidable problem in all its complexity. Preferences for various aspects

of the problem were expressed without losing the hope that the various approaches would ultimately furnish an integrated body of data. It appeared, however, that frequently individuals became committed to a goal and a method of research and seemed to overlook both the primary integrative aim and the limitations of their individual approaches.

In the course of the conference, strong needs to "proselytize" and to defend one's own position were evidenced. However, as the discussion progressed the participants resolved these tensions. This was accomplished either by explicit recognition of the importance of the work of others, or by free acceptance of the right of the individual to pursue his own interest.

The research described by the conference participants could be grouped according to three general areas: Outcome, Process, and Personality Theory.

Outcome.

Although great care was taken repeatedly to express the right of the investigator to conduct his research in any area and in any fashion that he deemed appropriate, the conferees implied that at the present time research in the area of outcome enjoys a status lower than that of either process or personality investigation. The tenor of the discussion strongly suggested that "outcome" research was generally scorned as being "applied," in contrast to the other two aims, which had the more lofty designation of "basic" research.

Although one of the prime purposes of the conference, as described in the original plan, was to provide "a comprehensive picture of the status of research on the *effects* of psychotherapy," as if by some tacit agreement the issue of outcome was skirted by the conference. When this fact was brought to the

attention of the membership near the end of the conference, various attempts were made to account for this phenomenon.

A. *Critics of outcome studies.* Many reservations were expressed by conferees regarding alleged characteristics and fallacies inherent in outcome research. The objections offered may be grouped as follows:

1. *Applied versus basic.* A recurrent theme was that to be concerned with outcome of therapy was to identify oneself with the simple pragmatic concerns of the practitioner rather than the more fundamental and more respected interests of the scientist. This differentiation carried a certain amount of admitted intellectual snobbery and a tendency to look down upon applied investigations. With a few notable exceptions researchers place high value on investigations which make extensions of theory. Outcome research was criticized for bearing no essential relationship to theories of personality, psychotherapy or psychology.

The attribution of such values to these research goals appeared to be based on the tacit assumption that new knowledge and creative insights are afforded by "pure research" while applied research merely seeks to employ knowledge already available. Further, pure research is a higher intellectual activity because it requires greater scientific ability and is more difficult. The validity of such arguments was doubted by those who pointed out that both pure and applied research are extremely difficult, require great skill and produce new knowledge.

2. *Non-specific effects of psychotherapy.* It was suggested that part of the reluctance to discuss outcome was due to the unexpressed fear that patient change may not be a consequence unique to psychotherapy. Also, unlike the model

of such wonder-drugs as penicillin and the mycin drugs which have specific effects, psychotherapy may have quite non-specific effects. It was implied that if analogies between drugs and psychotherapy were to be made, it might be appropriate to compare the influence of psychotherapy with the effects of that non-existent drug "bobe-mycin."² The responsiveness of some patients to psychotherapy is suspected to involve a strong element of suggestion. To confirm this would imply that the psychotherapist did not have a unique contribution to make in the treatment of patients. It was, therefore, too dangerous an area to study.

3. *Effectiveness of psychotherapy established.* In stark contrast to the misgivings expressed in the above view, others explained their reluctance to deal with outcome studies as based on the conviction that the effectiveness of psychotherapy had long since been established and need not be investigated further.

4. *Fear of guilt by association.* The quality of research that has been conducted and reported concerning outcome of therapy has in the opinion of many of the participants been impressively inadequate. Classically the area was treated by the loosest impressionistic reporting of individual successes and by the most superficial and unsophisticated research attempts. The effort may now be to dissociate oneself from this inadequate research product, lest the stigma of naïveté be applied indiscriminately to the unwary investigator.

5. *Psychotherapy is not an entity.* It was objected that research which focuses on the effects of psychotherapy makes the indefensible assumption that

psychotherapy is an homogeneous entity. A study of the effectiveness of an individual therapist's work with a specified patient is too gross since it does not differentiate what aspects of the treatment were therapeutic, psychonoxious or neutral. Instead such a study deals only with the end product, which may represent some complex response to the therapist's overall handling of the patient.

This error is also attributed to those who concern themselves with comparing the effects of different forms of therapy. Such research was described as neglecting the fact that the extent of variation in techniques by various therapists is probably as great within schools as between schools.

6. *Problems of selecting criteria.* Some critics of outcome studies considered the investigator's selection of specific criteria a premature and presumptuous value judgment. The selection of certain changes for study implies that a decision has been made regarding which alterations are therapeutic and which are not. Some argue that such evaluations are not in the province of the researcher and cannot as yet be supported. Others suggested that the selection of criteria must await further "process" studies.

A distinction was made between outcome research which was 1) the judgmental evaluation of whether a change is therapeutic or not, and 2) non-judgmental attention to change, independent of the valuation placed on it. This distinction lost some of its value when it was recognized that most investigators select for investigation those changes which they believe have some relevance to the therapeutic nature of the interchange. Changes are selected in accordance with some theoretical frame of reference and with some expectation that they are significant elements of the treatment process.

2. *Bobe-maysey* (Yiddish)—literally, grandmother-stories or old wives' tales.

Investigators often appear to have made their choices on the basis of their own predilections. They frequently make no effort to demonstrate the existence of a relationship between the psychological state studied at a given moment in time, and the subsequent course of the individual's behavior, in or out of the treatment setting. Frequently the relevance of the changes to the patient's psychodynamics are not made clear. Further, it is necessary to demonstrate that the personality changes measured are attributable to the unique function of the specific psychotherapy experience as distinguished from the restitutive effects of time or change in some unspecified conditions occurring within or outside of therapy. Another danger inherent in the specification of criteria of change is the temptation wittingly or unwittingly to conclude that conformity to cultural norms is evidence of improvement.

It was pointed out that most criteria of outcome are concerned with distress and disability but fail to deal with the positive aspects of mental health. Since mental health is not simply the absence of pathology, it is not sufficient merely to describe the melioration or elimination of symptoms. It is necessary also to study and define the healthy integrated individual. Much of the past and current study of behavior is limited to the pathological behavior side of the behavioral continuum.

B. Advocates of outcome studies. Although the majority of the conference members appeared to dissociate themselves from outcome studies there was a small group of supporters. One argument which they offered in support of "change" studies was based on the fact that when the therapist accepts a patient for treatment, he is implying that the patient is justified in his expectation that psychotherapy will be of benefit to him.

Although most of the conference participants preferred to identify themselves as researchers rather than practitioners, they recognized that to many therapists—and to society as a whole—outcome studies are of vital concern.

It was generally agreed that it would be most unfortunate for the general field of psychotherapy if research on outcome was avoided or allowed to languish simply on the basis of any or all of the objections listed above.

Process.

Perhaps the greatest interest was expressed in the conduct of studies dealing with the course of psychotherapy. The term "process," like most terms in the armamentarium of the investigator of psychotherapy, lacks definition. In the course of the conference the term process was applied with equal abandon to studies involving the analysis of moment by moment interchange between the patient and therapist in a given session and to studies utilizing general, abstract summary statements of the highlights of a prolonged period of psychotherapy. It appeared that anything less than a consideration of the "ultimate" change in patients subsequent to psychotherapy was lumped under the broad category of process analysis. Under process were classified both the description of changes which characterized a single patient's experience in the course of therapy, and the attempt to abstract similarities from the sequences of change experienced by groups of patients during their psychotherapy experience.

Although there appeared to be little or no objection regarding the wisdom of conducting process studies, some mildly sobering thoughts were expressed. Investigators were urged to consider the possibility that current hypotheses regarding the process of psychotherapy may have relevance only to a very small

proportion of the total patient population. This may be due to the fact that only a very small proportion of patients entering treatment are even "touched" by psychotherapy. This argument is closely related to the need to define criteria of change. If one wishes to study the process of psychotherapy then it is important to state with clarity and precision what one believes to be relevant. Clearly not all behavior which occurs in the treatment setting can be assumed to be therapeutic. A knowledge of the immediate, delayed, and cumulative effects of various interventions may be prerequisite for the pursuit of studies of the psychotherapy process. Can one study how a particular change comes about if that change is not designated in advance?

Personality Theory.

In some respects the study of basic personality variables is the most highly valued by all researchers. All investigators share the hope that their work may contribute to the basic store of knowledge and that meaningful general principles may be derived from their efforts. In this sense it may appear presumptuous to have a special category that endeavors to advance personality theory. Certainly the investigator who had identified his interests by the comparatively prosaic rubrics of outcome and process may feel he has been ployed by those who subscribe to the specific goal of advancing personality theory. The categorization is worth making, however, since the concern of such individuals with psychotherapy may be said to place primary emphasis on the "psycho" aspects of the word rather than on the "therapy." To them, the entire issue of the "therapeutic" or "psychonoxious" aspects of change is irrelevant. Such investigators maintained at the conference that the psychotherapy situation provides a laboratory particu-

larly appropriate to the study of the natural history of interactions among people. Insofar as the treatment setting prompts the patient to speak in an unusually frank and self-revealing fashion, not generally found in the context of general social interactions, psychotherapy offers unique opportunities for study. Such investigators may be able to utilize the data thus provided to amend theories of personality, learning, psychopathology and psychotherapy. To the degree that existing theories are improved with or without reference to the immediate clinical application of such concepts, the whole field may ultimately be benefited.

A good theory of personality would ultimately provide a good theory of psychotherapy. In the past, researchers have felt much less optimistic concerning the impact of their work on actual psychotherapeutic practice. However, at present it is felt that research may be effective in changing theory and therapeutic techniques.

As was characteristic of the handling of most issues on which there were divergences of opinions, the conferees agreed that in the light of the present inadequate state of knowledge of the field there was no advantage in attempting to persuade others to modify or change their interests; no single approach to the problems of psychotherapy appears to hold the greatest promise of success. Tolerance for the other researchers' interests and goals was stressed, with some emphasis being placed on the recognition of the natural divisions of research interest stemming from varying personal and professional backgrounds.

METHODS

The participants, in choosing methods appropriate to their various aims, appeared to fall readily into the two classical camps: 1) the experimenters, and

2) the observers or naturalists.³ Although there were "line crossers," as well as independents and straddlers, for purposes of characterizing the opinions expressed on methodological problems the dichotomy of groups may be useful.

The selection of a particular method for attacking a given problem appeared to be based not on its demonstrated unique appropriateness, but rather on the particular values and assumptions made by the investigator. As in the case of adherence to different goals, a member's identification with a particular methodology was often intense and ego-involved. The implication that one method was more effective or more appropriate was vigorously resisted by the proponents of the other method. The unkindest epithet that could be hurled at the other camp—and on one or two occasions it was placed in readiness on the launching pad—was "unscientific." Since this division of opinion appeared repeatedly to occupy the attention of the conference, it is appropriate to review here some of the differences in basic assumptions and values of each group. It may be more accurate to state that these differences represent values and assumptions *attributed* by members of one methodological school to members

of the other rather than values and assumptions actually held by the different schools. The description of these differences may caricature the positions of each group but, like any caricature, contains elements of fact.

Values and Assumptions.

A. *Identification with other sciences.* The experimental approach, which involves efforts to control and manipulate variables, appears to identify with the classical model of physics and chemistry. The effectiveness of such methods is evident in the phenomenal growth and advances of those fields. The observational approach appears to adopt as a model the astronomer, who attempts neither to manipulate nor to control but observes and orders his data and their relationships. Again, this model has proven its effectiveness by virtue of the accuracy of its predictions in astronomy. The observer in the field of psychotherapy may also model himself after Freud and other clinical giants in the field who have made their contributions to theory and practice on the basis of careful observation.

Advocates of each technique readily conceded, of course, that their methods, even if adhered to most rigorously, need not invariably lead to valid findings and conclusions. It was clear, however, that each thought that his model was more appropriate to the data treated and therefore more likely to achieve such valid conclusions.

B. *Complexity of variables.* The experimenter assumes that by reducing the number of variables he may gain knowledge; the naturalist assumes that by reducing the number of variables he inevitably loses knowledge. The experimenter has the conviction that although many variables may contribute to the total variance, only a small number, which he believes he has identified, will account for most of the variance. The

3. The boundaries of the two camps are not contiguous with those of the professions of psychology and psychiatry. In fact, the old stereotypes are rapidly becoming blurred. The very organization of the conference reveals a break with the traditional roles of the disciplines. The psychologist is usually characterized as being concerned with the rigor and vigor of the experimental approach, while the psychiatrist is viewed as being primarily concerned with the more subtle, sensitive analysis of the relationship between patient and therapist. It is interesting to note that in this conference, however, roles were reversed in that psychiatrists read formal papers on problems of experimental control while psychologists presented papers regarding patient-therapist relationships.

views of the opposing group of naturalist scientists may be classified as follows:

a) The extremists hold that the data of psychotherapy are inordinately complex and therefore any attempt to simplify them invariably introduces distortion. b) The moderates accept the principle that eventually a small number of variables may be identified as being the most relevant ones; however, they believe that designating these elements is at present premature.

C. Precision versus significance. The experimentalists are alleged to value the precise measurement of a variable, even a trivial one, in order to gain the satisfaction that comes from being sure of one's ground. This presumably satisfies the needs of an obsessional make-up. In contrast, the naturalists prefer to have even what is only a rough estimate of an important variable. It is postulated by some naturalists that a direct relationship exists between the triviality of a phenomenon and its ease of measurement. Data which lend themselves to simple direct measurement must, by the very nature of things, be unimportant and tangential. This implies yet another value that differentiates the two schools, namely, "tolerance for ambiguity." The experimentalist, presumably in deference to his need for structure, attempts to introduce order in a chaotic situation by selecting certain aspects of the total situation to study, and by dismissing from his attention all other data. The naturalist, however, is made less anxious by an ambiguous situation and is content to observe the naturally occurring event in its full complexity and postpone making formulations which may eventually resolve the ambiguity.

Research Practices.

The fear that manipulation and experimental control of variables might interfere with the normal development of the

process being studied produced quite different approaches to the practical problems of implementing research. The differences in approach were perhaps most clearly reflected in; data collection, rigors of design, and evaluation of evidence. There were also, of course, broad areas of agreement on each of these three issues regardless of points of view.

A. Data collection. Before deciding on the particular techniques of data collection to be employed, certain prior decisions must have been made. Investigators who believe that current knowledge in the field of psychotherapy is either grossly undependable or pitifully incomplete may feel impelled to begin their research *de novo*. These researchers are more often than not of the observationalist persuasion. Since the researcher may feel that no guidelines exist for determining which kinds of data are relevant and which are irrelevant he may feel impelled to select data gathering techniques that will represent as much of the total content as possible. Other investigators, generally experimentalists, appear to be more willing to risk specification of hypotheses and to gather only those data pertinent to such hypotheses. The observationalist, even when willing to designate the hypotheses underlying the study, prefers to include a much wider range of data in his studies in the expectation that this will enhance the opportunity for seeing new and unexpected relationships.

There was considerable agreement that in studying psychotherapy the researcher would prefer ideally to secure data regarding the treatment situation as perceived by each of the participants. The investigator may also wish to have an accurate representation of the reality or "true" situation as it transpired independent of how it was perceived. The problem for the investigator is whether to rely solely on the reports of "inde-

pendent" observers or to seek data directly from the participants. The experimenter may be inclined to include the therapy participants but the naturalist will generally not attempt to obtain reports from therapist or patient during the course of the therapy. The naturalist fears that his efforts to collect data may interfere with the natural process of the therapy he wishes to study. The naturalist's fears regarding interference with the therapy situation are greater with respect to collecting information from the patient than from the therapist since it is the usual practice for the therapist to keep notes regarding the interaction between self and the patient.

Members of both methodological schools perceived dangers inherent in the use of any observers, whether "independent" or "participant."

The following sections discuss some of the problems of data collection considered by the conference.

1. *Independent observers.* Some conferees stressed that the observer was a fallible "recording machine" in that he failed to secure all the data available and might selectively distort data on the basis of his own biases and internal stresses. Another view held that these risks were insignificant compared to the advantages gained by the use of a highly skilled, trained observer. Such an individual could, on the basis of his clinical experience and acumen, select and integrate the verbal and nonverbal cues in such a way as to make him a most sensitive instrument for collecting pertinent data.

The problem of how to reduce the error of an observer's report was not dealt with in detail. It was noted in passing that the expectation that validity of observations and conclusions could be enhanced by employing "expert" observers is not supported by experience. The reliability of observations and interpre-

tations between comparably expert persons is not impressive.

2. *Participant observer.* The use of the therapist as an observer poses a special set of problems. In addition to the usual dangers of the fallibility of the observer, it was argued that the therapist, perhaps by personality and personal stake in the theory and in the treatment, and certainly by his influence on the patient, is not in a good position to act as observer for research purposes.

The therapist by virtue of his helping role has a number of pressures on him. If one adds to this the requirement that he be reporter and observer his efficiency in performing each of these tasks may be reduced. The therapist occupies the unique position whereby his observations may be fed back to the patient, directly or indirectly. Thus he may influence the patient and therefore influence further the data in his observations. There is much evidence of the therapist's subtle impact on the patient in getting him to produce dreams, memories, etc., in conformity with the therapist's theories. It was recognized that this situation exists independent of whether the therapist acts as observer or not. The process of observing and reporting may simply intensify the therapist's need to support his reported observations. The therapist is, in effect, in a position to increase the likelihood of the "self-fulfilling prophecy."

3. *Mechanical recording devices.* Use of sound recording devices has become an accepted research procedure by almost all investigators. It appears to be a source of continual amazement and frequently of guilt-tinged gratification that the intervention of recording equipment is far less troublesome to the patient than the investigator generally had anticipated. The pliability of the patient, who appears willing to submit to a variety of techniques, is daily demonstrated both in therapy and research. The researcher

and therapist may not always match the patient's flexibility in this regard. Part of the patient's passive acquiescence may be due to the fact that wittingly or unwittingly the therapist and experimenter have communicated to the patient that such techniques are intended to be of benefit to the patient. This may explain both the patient's flexibility and the therapist's attendant anxiety.

4. *Research interventions.* Although the intervention of recording devices appeared to raise no special problem for the conference membership, specific objections were offered regarding other forms of experimental interventions. It was reasoned that unlike the inanimate recording device, the human observer-recorder or examiner may form relationships with the patient or offer material that may cause the patient to focus attention on certain aspects of his treatment contrary to the therapist's intent. The fear is not that the effect of the extraneous interventions will be non-therapeutic but simply that to the degree that they have any effect at all they are altering the normal course of therapy under investigation.

Inevitably the Heisenberg principle of indeterminacy was invoked in support of the notion that by the very act of measurement one distorts the data one is seeking to measure. Some of the participants took sharp issue with the seemingly popular practice of assuming that the Heisenberg principle as formulated regarding the quantum theory is appropriate for describing the approximate character of psychological knowledge. The principle of indeterminacy holds that operationally it is not possible to attempt an exact determination of the simultaneous velocity and position of an electron in its orbit in an atom because the very process of measuring the former inevitably and unpredictably alters the latter. The measurement problem as de-

scribed is inherent in the nature of the sub-atomic world as described by the quantum theory, and should not be generalized to large-scale events. It is clear, for example, that the simultaneous measurement of velocity and position of a planet or of a missile in its trajectory can indeed be measured with a high degree of accuracy.

In determining any single element of a large-scale psychological situation the researcher may so alter other aspects that he is frustrated in his attempt to understand the total situation. Such problems, however, are due to methodological difficulties and are not inherent in the psychological phenomena studied. Theoretically, refinement in measurement and method relating to large-scale phenomena is always possible and in principle attempts in that direction can constantly be made by the investigator. The approximate character of knowledge of large-scale events is related to measurement problems not unique to psychological knowledge but applying equally to any act of measurement.

The disagreement regarding the possible interference of data gathering techniques with the process to be studied is reflected in two contrasting views: a) since there is so little exact knowledge about the nature of the psychotherapeutic process, one need not be much concerned about inadvertently altering it by the techniques of study; b) since the boundaries of the therapy process are admittedly undefined, one must be particularly cautious not to interfere with the process by introducing any conditions that would artificially alter it.

5. *Psychophysiological measurement.* Particular attention was paid to the use of physiological measurement as a technique for studying psychotherapy. The observation that a relationship may exist between changes in an individual's affective state and changes in his physiolog-

ical reactivity has been the stimulus for a great deal of research. Since some changes occurring in the sympathetic nervous system can be measured with considerable precision, as contrasted to the gross estimates of the subject's psychological state, the potential contribution of physiological measures of psychotherapeutic process and outcome appears to be great. Discussions of the current status of such research noted that most of the reported investigations are guilty of one or more of three basic errors:

a. *Attributing observed changes in physiological measures directly to the effects of an hypothesized long-term distal psychological process.* Studies which cite changes in physiological responsiveness as evidence of the effects of a long-term psychotherapeutic process frequently fall into this error. Such studies fail to take into account that the autonomic response may be subject to short-term proximal transactional influences fully as much as is any other aspect of the patient's behavior and functioning. Convincing examples were offered to indicate that variations in an individual's physiological reactions to emotionally significant material may be grossly exaggerated or minimized by his perception of the quality of the therapist's "permissiveness and gentleness." Other studies have shown that responses may be extinguished independent of whether therapy is successful or not.

The researcher who wishes to use somatic change as a criterion of the effectiveness of psychotherapy received relatively little encouragement. He was reminded that many of the psychophysiological changes may be related to the testing or therapy setting to which the subject has adapted. It is necessary to show that the somatic effects are also demonstrable in other life situations. It is plausible, however, that the effective-

ness of the treatment of such illnesses as ulcer, headache, backache, hypertension, etc., should be reflected by changes in appropriate physiological measurements.

b. *Assuming that physiological measurement offers a direct representation of a complex psychological concept such as "arousal" or "affect."* Many studies appear to be motivated by the hope of finding physiological concomitants of the process of psychotherapy. Strong representations were made urging that researchers give up the goal of discerning a single index of sympathetic nervous activity which is directly comparable to a psychological construct, as the variables and their interrelationships are complex rather than simple.

Evidence was offered that a subject may have individual sensitivities to visceral reactions, and that people's modes of reacting to the same stimuli may vary considerably. Moreover, the same individual's reaction to the same stimuli may vary over time. The individual may show reactivity in one modality at one time and in a different modality at another time. The low intercorrelation among autonomic measures suggests that one measure cannot appropriately be substituted for another.

c. *Neglecting the influence of physiology on psychological state.* Much research emphasizes the influence of psychological factors on physiological phenomena but dismisses or underestimates the opposite relationship implicit in the James-Lange theory of emotion. Recent experiments have shown that patterns of autonomic response are associated with different states of receptivity to external stimuli. Such autonomic changes have been artificially induced by drugs in such a manner as either to enhance or to inhibit attention. Experimental evidence was also cited that some individuals are particularly alert to their own visceral changes. Such awareness

may initiate or reinforce an affective reaction. The physiological process may be viewed as one of the important communication systems that may operate both intra- and inter-personally.

It was modestly suggested by one who is well tutored in the field of psychophysiology that at this point the psychotherapeutic setting may have more to offer the field of physiology than has physiology to offer psychotherapy.

B. Rigors of design. The experimentalist, in deciding upon the techniques for the collection, analysis and reporting of data, takes cognizance of the standards and demands imposed by the tradition of experimental design. Philosophers of science have indicated some of the requirements which the experimenter must meet if the appropriateness of his conclusions is to be demonstrated or if evaluation and replication of the study by others is to be facilitated. Some naturalists appear to feel free from such constraints. Some conferees accused the observationalist of neither specifying nor being guided by any clear theory of naturalistic inference analogous to the existing system limiting experimental inference.

Both experimentalist and naturalist are in full agreement that the basic data of all research are the observations of the phenomena. Inferences are then drawn from these observations, which lead to hypotheses. One of the most distinguishing characteristics of the human mind is its capacity to produce a plethora of varied and ingenious "explanations" of the phenomena to which it selectively attends. Bavelas has recently found that it is almost impossible to present an individual with a set of data so random that the subject will be unable to form some conclusion regarding a systematic relationship among them. Perhaps even more disturbing is his finding that even when the subject is informed that only random

relationships exist, he remains adamant regarding the accuracy of his formulations.

Perhaps one of the major differences between the extremists of both schools lies in the degree of energy and attention devoted to the problem of rigorous and systematic testing of their inferences. The experimentalist is characterized by his careful efforts to control some variables and to vary others systematically in order to investigate and predict relationships among them. The naturalist may forego manipulation and control but retains the hope of making predictions on the basis of having accurately observed relationships in the data. Certain dangers and limitations of both positions when translated into practice were described

One of the assumptions of the experimental method is that the experiment is in fact a model of the event that occurs under natural conditions. The naturalists, as has already been described, generally believe that the complexities of psychotherapy do not lend themselves to experimental scrutiny. They further believe that it is not necessary to introduce controls to isolate the effect of various factors since the significant variables occur so powerfully in the natural situation that they can readily be identified. Moreover, replication of the power of a variable as it exists in nature may not be possible in the experimental situation. It was pointed out that experiments utilizing "normal" subjects rather than patients might lose considerable value in that the "attitudes and expectations" of patients toward the interventions of the therapist may not be replicated in a sample of normal subjects. The patient may generally come to the treatment situation with an expectancy of receiving help. The interventions of the therapist are perceived as potentially therapeutic. The subject involved in an experiment does not anticipate being benefited by

the experience but is oriented to wondering what the experimenter is up to. The responses of subject and patient to similar experimental procedures may, therefore, be quite different.

Another point questioned the ability of the experimental approach to meet its own requirements, namely, specifying the conditions with such clarity that others could subsequently replicate the work. The issue included the difficulty of specifying such pertinent variables as the patient's personality, salient features of the therapist's personality, the techniques of treatment employed in given situations, etc. Obviously the unique conditions that may obtain in one situation can never be reproduced in an absolute sense.

In defense of the experimental approach it was pointed out that the experimenter does not attempt to specify all conditions that are present but merely those which he hypothesizes are relevant to specified consequences. The experimenter is generally concerned with those aspects of the problem which permit generalization. To this end he may minimize the significance of individual uniqueness, and be tolerant of the fact that his techniques may not be suitable for the prediction of isolated and unique events that cannot be duplicated in another psychological situation.

Some naturalists emphasized the fact that their work involved prediction and was therefore quite systematic and rigorous. Moreover, the power of such predictions was considered great because the predictions were "complex." They attempted to spell out the contingencies under which certain events would be expected to occur. Such predictions would serve as a test of the relevance of the variables that had been designated for study. The discussants of this point cautioned that the use of predictions should pose some restraints on the pre-

dicator. The more intricate and conditional the contingency prediction becomes, the more difficult it is to derive quantitative functions, laws and regularities. The process of modifying predictions may continue until the prediction becomes overburdened with *ad hoc* additions and modifications. The point at which such a stage is reached is, of course, a matter of personal judgment and taste rather than cold logic.

C. Evaluation of evidence. It is obvious that the ultimate validity of an hypothesis or conclusion is independent of the method by which it was derived. From the viewpoint of advancement of the field the conceptualization that either makes or stimulates others to make salutory advances is welcomed without reference to its methodological ancestry. However, it was recognized that by virtue of his training, the experimentalist finds conclusions based on a tightly designed and rigorously executed investigation somehow more compelling. Similarly, findings based on careful and sensitive observation by individuals of recognized clinical acumen are more impressive to the naturalist.

The naturalist who presents hypotheses based on observations of a small sample without replication or controls, has in the view of the experimentalist left his work incomplete. While the naturalist may be content to leave the work of "testing and validating" his work to others, there are dangers in such a procedure. The individual testing hypotheses to which he has little or no commitment may often abandon them prematurely if his first attempts to confirm them are unsuccessful. Furthermore, the experimenter may fail to grasp the essential elements to be investigated and may use techniques too coarse or too refined for the problem. Presumably the individual who formulates the hypotheses would, if he were so inclined, be in a better posi-

tion to test his own formulations. The division between formulation and testing of an hypothesis has too often led to the unhappy situation where negative findings by the independent investigator are dismissed on the basis of his failure to comprehend the complexities of the situation. Support of one's hypotheses by another investigator is, of course, welcomed by the individual responsible for the initial formulation but too often shrugged off as simply establishing that which was already known.

Many investigators appeared to have little taste for submitting their hypotheses to experimental test. This may be based on the experience that offering one's carefully nurtured hypotheses to the scrutiny of colleagues may lead not simply to their examination but to the performance of a post-mortem as well.

All scientists make their obeisance to the notion that one must be loyal to the evidence. One must accept the ultimate supremacy of the facts over even the most cherished preconceived views. The fact remains, however, according to the views expressed at the conference, that only those conclusions are accepted by the investigator that "make sense" to him, i.e., are in accord with his previous experience and beliefs.

It was recommended that one procedure for evaluating the accuracy of one's formulations and constructs would be to subject them to independent verification outside of one's particular frame of reference. Such a test would involve finding important counterparts of the psychotherapy phenomena among the interpersonal relationships of the family, educational setting, etc. In support of this notion it was argued that it is extremely difficult for therapists or investigators to conceive of alternative ways of looking at the interaction that they are having with patients or subjects and of describing new orders of generalization other

than those they have been taught by their own experience. Nevertheless, some participants remained adamant, holding that it is quite irrelevant whether the concepts of psychotherapy find counterparts in other fields.

In the course of these discussions, the extreme positions were modified. It was generally agreed that in the field of psychotherapy, which is as yet "untilled empirically," the distinction between the experimental method and the naturalistic method of dealing with observational inferences disappears. Each approach reflects more the values and tastes of the researcher than any established authoritative evidence that one approach is more useful than the other.

SELECTION OF VARIABLES

Although there was considerable disagreement regarding goals and methods, there was general agreement that the following four major areas were appropriate concerns for research in psychotherapy: Form of Therapy and Techniques, The Therapist, The Patient, and Role of Theory.

Ideally the investigator might wish to comprehend the complex interaction among all variables. However, for the practical purposes of research, each of the variables is arbitrarily separated and studied. The particular element that one chooses to work with reflects one's own taste and clinical judgment.

Form of Therapy and Techniques.

The conference revealed relatively little interest in defending one "school" of therapy or attacking another. This non-committal attitude persisted despite the fact that one of the papers presented an extensive review of the effects of "leading" and "reflective" forms of therapy on specified groups of patients.

A serious question was raised, however, as to whether or not all participants

shared a comparable picture of the term psychotherapy. Were the investigators talking about the same concept when one based his discussion of psychotherapy on his experience with long-term intensive psychoanalyses while another was concerned with brief therapy situations exclusively? The problem was facetiously but effectively put by one of the participants who wondered whether the various investigators were perhaps not only blindly describing different parts of the proverbial elephant but were also palpating different kinds of beasts.

It was repeatedly pointed out that since there were such wide discrepancies among practitioners in their individual interpretations of any given form of therapy it was necessary to describe the particular style and patterning of techniques utilized by each therapist. The simple statement that the therapy was psychoanalytic or directive conveyed but little.

The Therapist.

This conference placed considerable emphasis on obtaining a fuller description of the therapist, who has generally been overlooked or treated only superficially in reports of psychotherapy studies. It was urged that in addition to describing techniques used, attention be paid specifically to the therapist's goals, personality, experience, and conviction.

A. *Goals.* For the investigator to deal intelligently with the data provided by the treatment situation, it is necessary that he understand the therapist's aims in such treatment. Such knowledge will guide the researcher in his selection of techniques for studying change and the process of effecting change. It is reasonable to expect that the therapist's aims will be related to the nature and direction of his interventions. Moreover, the therapist's own satisfaction with the course of therapy will depend on his

goals. Such satisfaction or the lack of it may play an important role in coloring his relationship to the patient

The therapist may seek to cure the patient or merely to ameliorate his "disability, distress and dread." It is also important to note whether the therapist wishes to help the patient achieve a greater measure of freedom or of conformity. One investigator reported he has found that the effective therapist is one who leads his patient in one or the other direction in a consistent fashion. His effectiveness with a patient does not depend on the nature of the goal *per se* but appears to be an interaction effect between the goals of freedom or conformity and the social class of the patient (See Patient section below.)

B. *Personality.* A number of discussants reported that a considerable part of the variance involved in therapeutic outcome is attributable to the personality of the therapist. Much of the literature regarding therapist's personality and attributes has focused on the negative influence of counter-transference rather than on the positive influences of the therapist's contribution. Such positive contributions are believed to result from the therapist's spontaneity, commitment, and efforts to understand the patient. Spontaneity refers to the degree to which the therapist appears free of constraints in relating to the patient. Commitment means the degree to which he conveys the attitude of trying to be helpful. With respect to efforts to understand how the patient perceives himself and the world, emphasis is placed not on the accuracy of understanding but on the ability to communicate interest in comprehending. These terms refer to the attitude of trying to be helpful which the therapist conveys to the patient. This involves not merely the therapist's acts but requires that the patient perceive the therapist as sincerely concerned with being useful.

The further point was made that commitment may arouse expectation of help in some patients but fail to do so in others. Expectations of help may be aroused in different patients by different kinds of behaviors on the part of the therapist; when aroused, such expectations generally contribute to the therapeutic experience of the patient.

The therapist's attitude toward the patient appears to be related to the ultimate success or failure of the course of therapy. For example, high regard for the patient is reflected in success while antipathy appears to be correlated with failure. Presumably the therapist's attitude is instrumental in effecting the observed changes in the patient; however, the possibility cannot be ruled out that the therapist's attitude may at least in part be in response to the patient's reactions to his efforts.

A warning was sounded against the practice of categorizing therapists' interventions as "good or bad," in that patients may respond variously to the same acts. The effectiveness of an intervention by the therapist depends on its meaning to the patient.

According to the various comments of the participants, a meaningful description of the therapist may require the inclusion of such information as the therapist's perception of the patient, his preference for a given type of therapy, his motivation, his expectations and goals, his self-insight, insight into others, adaptability, emotional control, dependence, degree of ambiguity in the cues he presents to the patient, etc. Each of these factors may subtly influence the therapist's behavior toward the patient and the patient's response to him.

C. Experience. The point was frequently made that if one wishes to investigate psychotherapy, every attempt should be made to utilize the work of the expert therapist rather than the

novice. Presumably the expert therapist would afford more opportunity to the researcher to observe therapeutic interchanges and events.

A similar argument for the use of fully qualified therapists was made by those who expressed concern regarding the issue of responsibility to the patient. The ethical question, implicit in entrusting the treatment of disturbed patients to the relatively untutored therapist, was briefly explored.

On the other hand, it was pointed out, there is no clear evidence to date to show that the inexperienced therapist is any less effective or more damaging than the experienced therapist. It was also stated that the less experienced therapist frequently achieves even greater success than the more experienced one. This disturbingly perverse finding was supported by a number of the participants. Many experienced therapists are of the opinion that as novices they were able to effect dramatically therapeutic changes in patients which they frequently were unable to repeat after having achieved greater professional maturity. One of the participants was sufficiently intrigued by this alleged fact to propose half-seriously the provocative hypothesis that the more experienced therapists are the more probing, directive therapists and may by these procedures be producing the "sickness" (distress) in the patient. It was further noted that, since there is little agreement in the field regarding standards by which to assess the effectiveness of psychotherapy, the question of the relative merits of therapy as performed by experts or novices cannot be meaningfully discussed.

Some members of the medical profession held that quite independent of the question of experience and competence, the practice of psychotherapy should be limited to the M.D. Little or no effort was made to debate the question of the

guild rights of one profession over the other. Instead the focus remained on the problem of assessing the qualifications *per se* of an individual for conducting

D. *Conviction.* One of the characteristics of the therapist that was frequently asserted to be related to his effectiveness was his confidence in the efficacy of the techniques and theories on which he based his therapy. The specific dynamics that operate to produce this phenomenon are not clear. It may be that the therapist's conviction serves to reduce his own anxiety to the point where he can more effectively understand the patient's communications; or his confidence is communicated to the patient and serves to reassure the patient and thereby enables him to be more accessible to the help offered by the therapist. In any event, the therapist's assurance is believed to be a significant element of the psychotherapy situation. The question was then raised about the appropriateness of studies which require the therapist to treat patients by techniques to which he does not wholeheartedly subscribe. If one were to require a therapist experienced in a form of therapy like the psychoanalytic to utilize the non-directive approach, at least two sources of error might be introduced into the experiment: 1) the therapist's lack of conviction, and 2) his lack of familiarity with and skill in handling the new techniques.

Frequently studies which attempt to compare the effects of two or more forms of therapy with comparable patients make use of relatively inexperienced therapists. It may be argued that the factors of experience and conviction may be held fairly constant under such conditions by virtue of their absence rather than by their presence. On the other hand, to many, testing a form of therapy as practiced by novices rather than experts seemed a dubious procedure.

The Patient.

In contrast to the majority who laid heavy emphasis on the role of the therapist, there were those who stressed the significance of the patient variable independent of the therapist's function. One investigator reported that practically all the variance in outcome was found to be a function of variance within the patient at the beginning of therapy. He believed that only in the most extreme cases could one expect that success or failure in therapy would be a function of the therapist's personality in relation to his patient. A further point regarding the role of the patients was made concerning goals in therapy. It was hypothesized that the goal of inner freedom is more valued by intellectuals and middle-class persons than by lower-class persons. The contrast between the therapist's characteristic middle-class value on freedom and the lower-class patient's desire for structure may be a significant factor in explaining the high rate of drop-outs found among the lower socio-economic groups. Thus, patients may be expected to differ in their susceptibility to the therapist's leadership in the direction of either conformity or freedom.

Role of Theory.

Although research is often construed as a search for certainty, it was recognized that one of the characteristics of science is that it invites questioning. It grows because no one proposition is in itself absolutely certain. The process of correction operates when adequate evidence is found to test the "truths of science." In contrast, a system ceases to be scientific when it cannot admit that it may be mistaken in any part of its system. Too often researchers affiliated with one or another theoretical framework have fallen victim to the malady of premature hardening of the categories. As a result many beliefs are viewed as confirmed, and unnecessary of further

scrutiny, while other concepts are accepted as essentially untestable.

It is recognized that dangers exist in adopting either the position of the chronically "open mind," or an obstinate irrevocable attachment to a particular theory. The investigator must take a theoretical stand but be prepared to modify his views in accordance with new evidence. As in all science, no great achievements are possible without courage.

Researchers in the field of psychotherapy have long nurtured the belief that research might best be advanced by encouraging investigators to utilize their individual approaches and their particular frames of reference. It was anticipated that such independent approaches might eventually result in the discovery and rediscovery of facts that would merge to permit the development of an improved theory. This expectation has

not yet been fulfilled. At least one participant voiced the view that this *laissez-faire* approach has "played itself out." Although there has been a great accumulation of clinical observations and experimental findings, the field has made relatively little progress. There has been little creative building on the work of others. This is due in large measure to the fact that current reports lack the clarity and completeness which are necessary to permit careful analysis, comparison, replication and stimulation of further research. It was suggested that perhaps now researchers might be willing to define more explicitly their goals, methods and variables selected. In short, more attention should be paid to enhancing comparability of independent works. Hopefully this would facilitate the growth of a systematic body of knowledge prerequisite to the development of the science of psychology.

EPILOGUE

Almost a decade ago the Conference on Graduate Education in Clinical Psychology facetiously characterized the status of the field of psychotherapy as follows: "Psychotherapy is an undefined technique applied to unspecified problems with unpredictable outcome. For this technique we recommend rigorous training." Much research has been done since that comment was made, but there has been relatively little progress in establishing a firm and substantial body of evidence to support very many research hypotheses.

Basic problems in this field of research have remained essentially unchanged and unsolved. This may be due in large part to the fact that both the investigator and

the therapist have managed to preserve their favorite concepts, assumptions, values and hypotheses by hermetically sealing them in layers of ambiguity. This conference was successful to the degree that some of this ambiguity was recognized if not actually pierced.

Participants came to this conference prepared to defend their own views and to demonstrate the inadequacies of the views of others. In the course of the meeting it became apparent that there is no simple, reassuring, authoritative principle which clearly supports one approach and demonstrates the invalidity of the others. Ultimately each investigator must fall back on his own values and tastes. Remarkably little effort was devoted to proselytizing for the acceptance of the

superiority of one's own goals, methods and techniques. Instead, considerable interest was expressed in determining how individuals of intelligence, experience and integrity arrived at and maintained their contrary views. As a consequence this curiosity led the membership to take the first step in initiating meaningful interchange. Investigators were able to listen to each other and to experience to a surprising degree the development of mutual respect. Under conditions of

such respect communication was remarkably enhanced

The experience of this conference appeared to confirm the views of Vannevar Bush who recently commented, "Men do not learn to understand one another merely by sharing intellectual experiences. They must meet on an emotional level if the foundation is to be built for collaboration on a high plane." This conference provided both the intellectual and emotional experiences in full measure.

UNIVERSAL
LIBRARY



134 876

UNIVERSAL
LIBRARY